

ILLEGITIMATE RECOMBINATION IN PLASMIDS

DAVID PINDER

**Thesis presented for the degree of Doctor of Philosophy
University of Edinburgh**

1996



Abstract

Illegitimate recombination mechanisms are important for genetic change within an organism. They are also the cause of many instability problems in biotechnology and have been associated with certain human genetic disorders and cancers.

The original aim of this work was to construct a deletion (illegitimate recombination) resistant cosmid based cloning system, for the cloning of unstable human DNA. Two 'mutant plasmids' pMS5 and pMS7 were isolated. The plasmids were derived from pUC18 and appeared to stabilise the propagation of a long DNA palindrome. The basic concept was to construct a new cosmid using pMS7 as part of the backbone.

The construction of two new cosmids cDRII (Deletion Resistant) and cDRIII is described. However, they are unlikely to contain a mutation which stabilises unstable sequences.

This thesis also describes the search for the 'mutation' in pMS7 by, single-strand conformation polymorphism and fragment swap analysis. This work did not detect any such mutation. However, it lead to the isolation of a novel mutation composed of both direct and inverted repeats, which I have called DIR. The presence of DIR in pAC2 (a derivative of pUC18, with the same DNA palindrome as pMS7), and the absence of DIR from pMS7, are consistent with the failure to find a stabilising 'mutation' in pMS7. The structure of the DIR mutation is analysed in detail and a hypothesis for its formation is proposed.

Finally, the behaviour of four long DNA palindromes (other than that cloned in pMS7) is investigated, when ligated into pM* (a palindrome-free derivative of pMS7).

Acknowledgements

I would like to thank David Leach for scientific guidance, Thorsten Al lers, Catherine Blake, John Connelly, Angus Davison, Phil Eastlake, Ewa Okely and Noreen Murray for helpful discussions and technical advice, and NBL Gene Sciences Limited. Lastly I wish to thank my wife, Sylvia, for her patience, support and understanding.

Abbreviations

Amp ^r	Ampicillin resistant
AMPS	Ammonium persulphate
ATP	Adenosine-5`-triphosphate
BAC	Bacterial Artificial Chromosome
bp	Base pair(s)
BIME	Bacterial Interspersed Mosaic Elements
BSA	Bovine serum albumin
°C	Degrees Centigrade
cm	Centimetre(s)
<i>cos</i>	Cohesive λ termini
(d)dATP	2` (3`-di)-deoxyadenosine-5`-triphosphate
(d)dCTP	2` (3`-di)-deoxycytidine-5`-triphosphate
(d)dGTP	2` (3`-di)-deoxyguanosine-5`-triphosphate
(d)dTTP	2` (3`-di)-deoxythymidine-5`-triphosphate
(d)dNTP	2` (3`-di)-deoxynucleoside-5`-triphosphate
DIR	Direct and Inverted Repeat
DNase	Deoxyribonuclease
DTT	Dithiothreitol
EDTA	Diaminoethanetetra-acetic acid
ERIC	Enterobacterial Repetitive Intergenic Consensus
g	Gram(mes)
HEPES	N-2-hydroxyethylpiperazine
IHF	Integration Host Factor
in	Inch(es)
IPTG	Isopropylthio- β -D-galactoside
IRU	Intergenic Repeat Unit

IS	Insertion Sequence
Kan ^r	Kanamycin resistant
krpm	Kilorevolutions per minute
kb	Kilobase pairs
λ	Bacteriophage Lambda
l	Litre(s)
lb	Pound(s)
MAC	Mammalian Artificial Chromosome
mg	Milligram(s)
ml	Millilitre(s)
M	Molar
mm	Millimetre
mM	Millimolar
mol	Mole(s)
μCi	Microcurie(s)
μg	Microgram(s)
μl	Microlitre(s)
<i>ori</i>	Origin of replication
PAC	P1-derived Artificial Chromosome
pH	$-\log_{10}[\text{H}^+]$
REP	Repetitive Extragenic Palindromes
RIP	Repetitive IHF-binding Protein
RIB	Reiterative IHF BIME
RNAse	Ribonuclease
SDS	Sodium dodecyl sulphate
SSCP	Single-Strand Conformation Polymorphism
TEMED	N-N-N'-N'-tetra-methyl-1,2-diamino-ethane

Tris	2-amino-2-(hydroxymethyl)-1,3-propandiol
Triton X-100	Octylphenooxypolyethoxyethanol
U	Unit(s)
V	Volt(s)
v/v	Volume per volume
v/w	Volume per weight
W	Watt(s)
w/v	Weight per volume
Xgal	5-bromo-4-chloro-3-indoly- β -D-galactoside
YAC	Yeast Artificial Chromosome

Table of Contents

TITLE PAGE	i
DECLARATION	ii
ABSTRACT	iii
ACKNOWLEDGEMENTS	iv
ABBREVIATIONS	v
TABLE OF CONTENTS	viii
CHAPTER 1:.....	1
INTRODUCTION	2
SECTION 1:.....	4
<i>Genetic Recombination</i>	6
Homologous Recombination.....	6
Site-Specific Recombination.....	7
Transposition	7
<i>Illegitimate Recombination</i>	7
Classes of Illegitimate Recombination.....	8
Strand Slippage.....	8
Strand Slippage and Secondary Structure.....	9
Strand Slippage associated with <i>E. coli</i> Multicopy Plasmids	9
Strand Slippage and Disease.....	15
<i>DNA Palindromes</i>	20
Possible Cellular Roles for Palindromic DNA.....	20
Evolution	20
Transcription & Replication	22
Mobile Genetic Elements.....	22
Repetitive Palindromic DNA.....	22
Human Genetic Disease.....	23
<i>Long DNA Palindromes in Escherichia coli</i>	23
Palindrome Length.....	24
Sequence of Palindrome.....	24
Short Direct Repeats	25
Genotype of Host.....	26
DNA Replication.....	26
SECTION 2:.....	28
<i>Cosmid Cloning</i>	28

<i>Introduction</i>	29
<i>Cosmid Vectors</i>	29
<i>In vitro</i> Packaging of Cosmid DNA.....	30
The Development of Cosmids.....	32
Problems Associated with the use of Cosmids.....	33
<i>Yeast Artificial Chromosomes</i>	35
Problems Associated with the use of YACs.....	35
<i>Summary</i>	35
CHAPTER 2:	37
MATERIALS & METHODS	37
MATERIALS.....	38
<i>Microbiological Strains, Media and Solutions</i>	38
Strains	38
Media	38
Bacteriological Media.....	38
Media Additives.....	38
Materials for DNA Purification and Manipulation	39
General Solutions and Materials for DNA Purification and Manipulation.....	39
Solutions for Ethanol Precipitation.....	39
Materials for Purification of DNA from Agarose Gels and Solutions	40
Solutions for Plasmid/Cosmid DNA Purification's	40
Rapid Plasmid Preparation.....	40
QIAGEN Plasmid Purification	40
SDS Alkaline Lysis Plasmid DNA Purification.....	40
Triton Lysis Midi Plasmid DNA Purification.....	40
Preparation of Cosmid DNA from Minilysates	41
Enzymes, Buffers and Solutions for DNA Manipulation.....	41
Restriction Enzymes	41
DNA Sequencing.....	41
Other Enzymes.....	41
Other Solutions	41
Solutions for Gel Electrophoresis	41
Agarose Gel Electrophoresis.....	41
Polyacrylamide Gel Electrophoresis.....	41
Solutions for Silver Staining.....	43

Solutions for Cosmid Methods	43
<i>In Vitro</i> Packaging of Cosmid DNA	43
Transformation	43
Electroporation	43
METHODS	46
<i>DNA Purification and Manipulation</i>	46
Ethanol Precipitation.....	46
Purification of DNA in Agarose Gels or Solutions.....	46
GENECLEAN® II Kit.....	46
QIAquick™ Gel Extraction Kit	47
<i>Plasmid/Cosmid DNA Purification</i>	48
Rapid Plasmid Preparation.....	48
QIAGEN Plasmid Purification Method	48
SDS Alkaline Lysis Plasmid DNA Purification	49
Triton Lysis Midi Plasmid DNA Purification	50
Preparation of Cosmid DNA from Minilysates.....	51
<i>General DNA Manipulation</i>	51
Annealing of Oligonucleotides	51
Restriction Enzyme Digests	51
DNA Sequencing	52
DNA Ligation	52
DNA Dephosphorylation	53
Radiolabelling of DNA	53
<i>Gel Electrophoresis</i>	53
Agarose Gel Electrophoresis.....	53
Polyacrylamide Gel Electrophoresis	54
Sequencing Gels	54
10% Longranger™ Gels	55
SSCP Gel Electrophoresis	55
Autoradiography	55
<i>Microbiological Methods</i>	56
Bacterial Methods	56
Storage of Bacteria.....	56
Growth of Bacteria.....	56
<i>Cosmid Cloning Methods</i>	56
<i>In Vitro</i> Packaging of Cosmid DNA	56

<i>Plasmid Methods</i>	57
Maintenance of Plasmids	57
Transformation.....	57
Electroporation.....	57
CHAPTER 3:	59
<i>The Search for the 'Mutation' in pMS7</i>	59
<i>Introduction</i>	61
<i>Results</i>	62
SSCP analysis	62
Fragment Swap Experiments	63
<i>Conclusion</i>	71
SSCP analysis	71
Fragment Swap Experiments	71
CHAPTER 4:	73
<i>The cloning of DNA palindromes</i>	73
<i>Introduction</i>	74
<i>Results</i>	75
Cloning Other DNA Palindromes	75
Lowered DNA Yield/Copy Number	76
<i>Conclusion</i>	80
Cloning Other DNA Palindromes	80
Lowered DNA Yield/Copy Number	80
CHAPTER 5:	82
<i>The isolation of a novel mutation: DIR</i>	82
<i>Introduction</i>	83
<i>Results</i>	84
Cut and Ligate experiment.....	84
Fine-Structure Analysis of pDIR1	85
Sequencing of the 42bp Insert Present in pDIR1	85
Analysis of pDIR1 Derivatives	86
<i>Conclusion</i>	95
CHAPTER 6:	97
<i>The Construction of cDR Cloning Vectors</i>	97

<i>Introduction</i>	98
<i>The Construction of cDRI</i>	101
<i>The Construction of cDRII & cDRIII</i>	101
a) Summary of the structure of <i>cos</i> DNA in λ and the 431bp <i>MboI</i> λ fragment.....	102
b) Isolation of the <i>cos</i> Containing fragment from Lambda.....	102
c) Ligation of Oligo1 into pM*	103
d) Ligation of <i>MboI</i> λ Fragment into Oligo1, in pM*	103
e) Ligation of Oligo2 into Oligo1, in pM*	103
f) Ligation of <i>MboI</i> λ Fragment into Oligo2	104
g) Isolation of the Kanamycin GenBlock	104
h) Ligation of Kanamycin GenBlock into <i>BglII</i> site in Oligo1.....	104
i) Scheme of Use for cDRII & cDRIII.....	105
<i>The cloning efficiency of the cDR cosmids</i>	105
<i>Results</i>	106
<i>Conclusion</i>	117
CHAPTER 7:	118
DISCUSSION	118
<i>Summary</i>	119
<i>Discussion</i>	121
Model for the Formation of DIR.....	121
The future Development of cDRII and cDRIII	124
BIBLIOGRAPHY	125
APPENDIX 1	148

CHAPTER 1:

INTRODUCTION

INTRODUCTION

The survival of a species depends on its ability to maintain genetic diversity, so that individuals can vary their response to unforeseen environmental changes. Diversity is maintained through mutation, which alters single genes or small groups of genes in an individual, and recombination, which rearranges genetic material.

Genetic rearrangement may be mediated by many mechanisms including: transposition, homologous recombination, site-specific recombination and illegitimate recombination (Leach, 1996). Illegitimate recombination is a mixture of events and cellular mechanisms (for a recent review see Ehrlich, 1993), but may be divided into two main classes, end joining and strand slippage. When cloning eukaryotic DNA, instability is frequent among cloned sequences. Much of the observed instability is due to illegitimate recombination.

At the onset of this work, the goal was to construct a “deletion resistant” (illegitimate recombination resistant) cosmid based system, for the cloning of unstable human DNA sequences. Two ‘mutant’ plasmids, pMS5 and pMS7 were isolated. These plasmids were derived from pUC18 and appeared to stabilise a 571 base pair (bp) palindrome. The basic concept was to use pMS7 as part of the backbone in the construction of a new cosmid. This cosmid would then contain the ‘mutation’ and would have the potential to stabilise unstable sequences. The 571bp palindrome was cloned into the lacZ’ gene of pUC18, and instability of the palindrome was estimated by the phenotype of clones on special media. After an extensive search for the ‘mutation’ in pMS7 an alternative explanation was discovered. The parent vector pAC2 from which pMS7 was derived was found to contain an unusual mutation which gave the appearance of instability. I have called this mutation DIR for direct and inverted repeat.

This thesis describes: a) the search for the ‘intangible mutation’ in pMS7 by:
1) Single-Strand Conformation Polymorphism (SSCP) analysis comparing pM*

(pMS7 derivative, without palindrome) and pUC18. 2) fragment swap analysis where fragments from various plasmids were swapped, to determine the fragments which generated a change in behaviour, b) the behaviour of different palindromes when cloned in pM*, c) the isolation of a novel mutation composed of both direct and inverted repeats, that I have called DIR (Direct and Inverted Repeat), referred to as the “DIR mutation” in this thesis (the data support the hypothesis that DIR is responsible for the differences of plating behaviour in pMS7 and pAC2), and d) the construction of three cosmid vectors. The results describing the discovery of DIR have been submitted for publication (Pinder *et al.*).

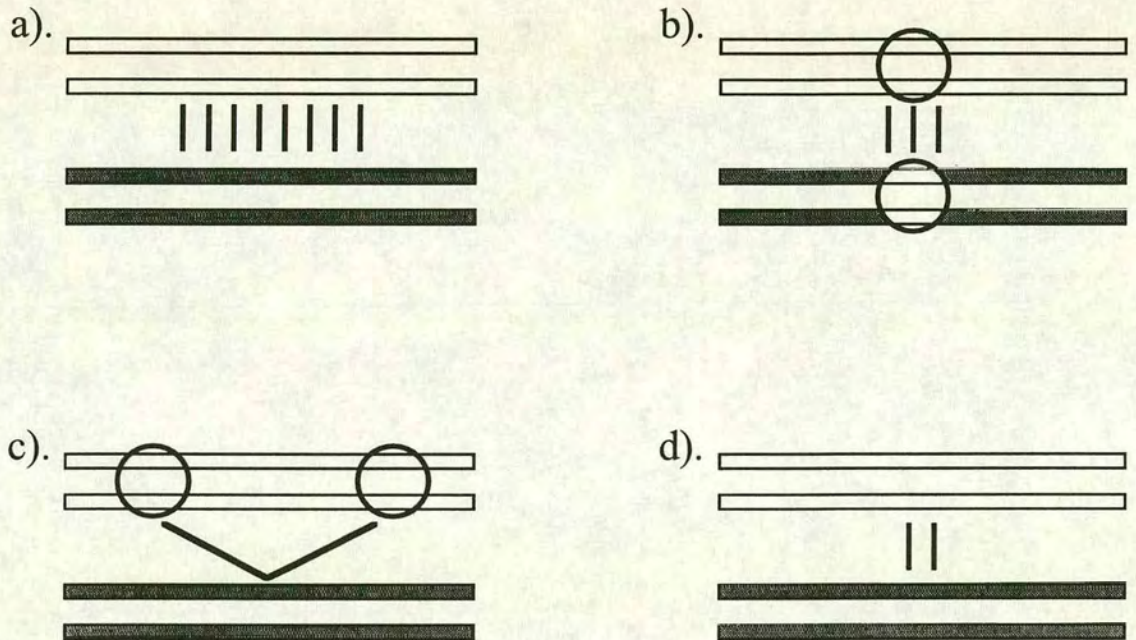
This chapter is split into 2 sections, the first is an introduction to recombination, illegitimate recombination, strand slippage, DNA palindromes and other topics related to this thesis; the second is an introduction to cosmids and yeast artificial chromosomes, their uses as cloning vectors and associated problems.

SECTION 1:

Illegitimate Recombination

Figure 1.1

Overview of the four classes of recombination (Leach, 1996): a) homologous recombination b) site-specific recombination c) transposition & d) illegitimate recombination.



Genetic Recombination

Genomes are not static, they have both the necessary levels of fidelity for stable inheritance, and the required flexibility for genetic change. The survival of, or the ability for a species to evolve, depends on genetic change. This is primarily mediated in genomes by both mutation and recombination. Genetic recombination in both prokaryotes and eukaryotes, may result from a). 'homologous recombination', recombination between long homologous sequences, b). 'site-specific recombination', mediated by specialised proteins, acting as catalysts, that bind specific target sequences, c). 'transposition' (of transposable elements), mediated by proteins that recognise the ends of the element, d). and from 'illegitimate recombination', between sequences of little or no homology (see Fig 1.1 for overview of recombination mechanisms). As this thesis is associated mainly with illegitimate recombination, it will be discussed in detail in the following sections.

Homologous Recombination

The roles of homologous recombination in cells are varied. The most important of these is the maintenance of genetic diversity. Other consequences of homologous recombination include: a). repair of double strand DNA damage, b). scheduled DNA rearrangements such as mating-type switching in fungi or antigenic variation as seen in various pathogens, and c). the spread of selfish DNA sequences as seen with mobile introns. Homologous recombination in prokaryotes has been studied most in the Red and Rec pathways. In *Escherichia coli* homologous recombination is RecA-dependent (reviewed by Leach, 1996; Kowalczykowski *et al.*, 1995; Smith, 1988; West 1992).

Site-Specific Recombination

Site-specific recombination involves the interaction of defined DNA sequences. Unlike homologous recombination, the proteins involved in site-specific recombination mediate recognition of specific sequences directly. The best studied models of this type of recombination are: a). bacteriophage λ integration/excision and b). Tn3 co-integrate resolution (reviewed by Leach, 1996; Sadowski, 1993).

Transposition

A transposable element may be defined as a DNA sequence capable of inserting itself into many different sites within the host's genome. Most transposons share the common feature of inverted terminal repeats. A transposase recognises the ends of the cognate element, breaks the DNA at these ends, and joins the ends to target sequences (reviewed by, Leach, 1996; Berg & Howe, 1989). Some transposons may take part in illegitimate recombination events also.

Illegitimate Recombination

The phrase illegitimate recombination was originally used by Franklin (1971), to define recombination between sequences with little or no homology.

Illegitimate recombination mechanisms are important in evolution (Anderson & Roth, 1977; Ohno, 1970). During the process of evolution large genome rearrangements, such as duplications, deletions, translocations and insertions can be extremely advantageous. In the case of genetic material which has been duplicated, this provides supplementary copies of genes, which can accommodate mutations and thus evolve. Translocations and deletions can fuse genes and create proteins with novel function, or simply alter the regulatory environment of a gene. Insertion of foreign DNA into a genome facilitates horizontal transfer, allowing different organisms to bypass the need to repeatedly evolve similar functions (Campbell,

1979; Arber, 1984). In biotechnology, 'the utilisation of living organisms for the benefit of man', illegitimate recombination is frequently a problem, and *in vitro* constructed genomes are often unstable (Ehrlich *et al.*, 1986; Primrose, 1981; Summers, 1994). Illegitimate recombination has also been found associated with certain human genetic disorders (reviewed by Richards & Sutherland, 1992; Wells and Sinden, 1993; Wells, 1996), and colorectal carcinomas (Thibodeau *et al.*, 1993; Ionov *et al.*, 1993).

Classes of Illegitimate Recombination

Illegitimate recombination may be divided into two broad classes: a). end joining, which is mediated by enzymes which cut and join DNA, such as topoisomerases, site-specific DNases and proteins which initiate rolling circle replication, b). strand slippage, where, after pausing at the replication fork, the nascent strand can dissociate from one template and pair with another (reviewed by Allgood & Silhavy, 1988; Ehrlich, 1989; Franklin, 1971; Leach, 1996; Meuth, 1989). For the purposes of this thesis strand slippage will be discussed only.

Strand Slippage

Strand slippage is thought to be mediated by a pause in DNA replication (for a review see Bierne & Michel, 1994). Several structural and functional features can regulate the rate at which the replication fork progresses. These include DNA secondary structures, DNA damage, and structural or regulatory proteins tightly bound to the DNA.

Strand slippage in *E. coli*, usually occurs in association with short homologous sequences, but unlike homologous recombination this process is *recA*-independent both in plasmids (Matfield *et al.*, 1985; Lovett *et al.*, 1993; Bi & Liu, 1994), and the chromosome (Lovett *et al.*, 1993). A review by Kuzminov (1996) concludes that inhibition of the progress of replication forks in *E. coli* causes their instability.

Strand Slippage and Secondary Structure

Sequences which have the potential to form unusual secondary structure are known to promote pausing of DNA replication and can stimulate replication slippage. Such sequences include: a) oligopurine-oligopyrimidine sequences, able to adopt a triple helix structure in supercoiled DNA (H-DNA), both *in vivo* (Rao *et al.*, 1988) and *in vitro* (Dayn *et al.*, 1992; Samadashwily *et al.*, 1993), b) tri- and tetra-nucleotide repeats, thought to form pseudo-hairpins when single stranded (Wells, 1996; Richards & Sutherland, 1992), and recently shown to have the potential to form a K⁺-dependent quadruplex secondary structure (Usdin & Woodford, 1995), c) alternating GC or GT dinucleotide repeats (microsatellite repeats), capable of forming Z-DNA (Hite *et al.*, 1996; Freund *et al.*, 1989; Klysik *et al.*, 1982), and d) DNA palindromes (perfect inverted repeats containing no spacer region at the central sequence), able to form hairpins when single stranded, or cruciforms when double stranded (for a review see Leach, 1994). A diagrammatic representation of possible DNA secondary structures, known to stimulate replication slippage, may be seen in Fig.1.2.

Strand Slippage associated with *E. coli* Multicopy Plasmids

Replication slippage events in *E. coli* multicopy plasmids are: a) usually mediated by short (~3-20bp) homologous sequences (Bi & Liu, 1994; Bi & Liu, 1996), b) mostly, *recA*-independent (Allgood & Silhavy, 1988; Bi & Liu, 1994; Bi & Liu, 1996; Lovett *et al.*, 1993), c) sharply reduced by increasing the distance between repeats (Lovett *et al.*, 1993; Chédin *et al.*, 1994; Bi & Liu, 1994; Bi & Liu, 1996), and d) produce multiple products (Dianov *et al.*, 1991; Mazin *et al.*, 1991; Lovett *et al.*, 1993; Bi & Liu, 1994; Bi *et al.*, 1995; Bi & Liu, 1996).

An emerging theme of experiments using multicopy plasmids, in the study of recombination and mutagenesis, is that deletion formation is always associated with

Figure 1.2

Secondary structures potentially formed by trinucleotide repeats and oligopurine-oligopyrimidine sequences: i) model for triple helical structure (H-DNA) formed by oligopurine-oligopyrimidine stretches (Hoogsteen base pairing is represented by four pointed star), ii) models for the K^+ -independent (a) and K^+ -dependent (b) structures formed by $(CGG)_{20}$ tract. Tetrads are shown as grey parallelograms. Different loop sizes, loop orientations and tetrad numbers are possible (Usdin & Woodford, 1995).

Figure 1.2i

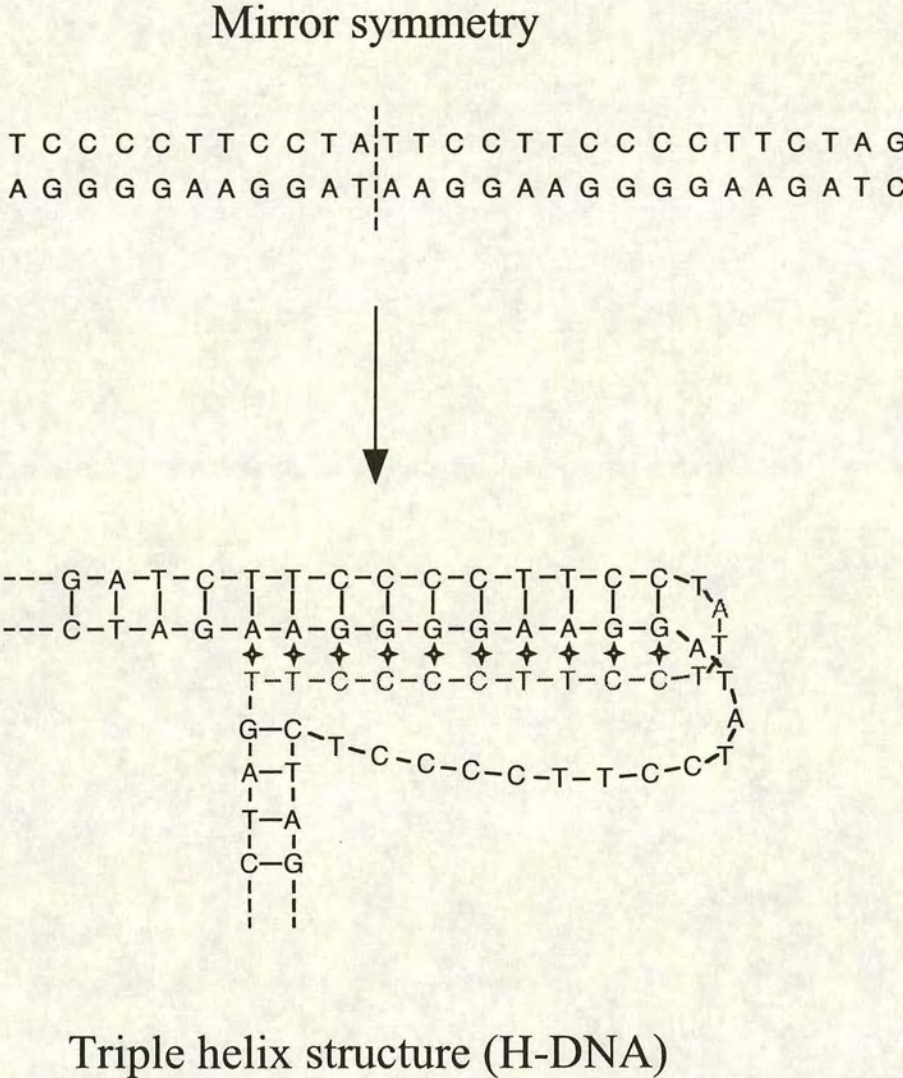
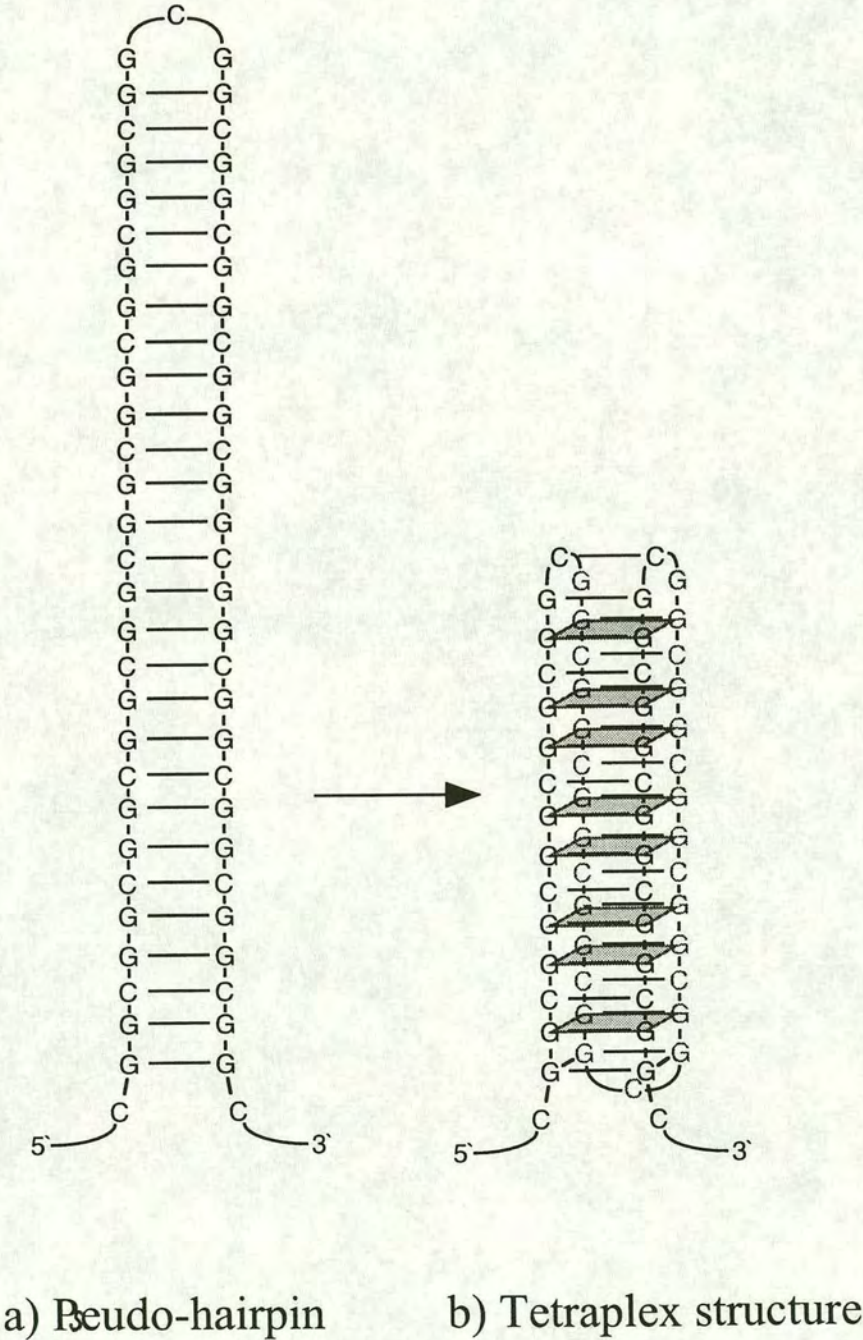


Figure 1.2ii



dimerization of plasmid DNA (for a recent review see Kuzminov, 1996). This suggests an intimate involvement of recombination in the deletion process. More than one mechanism may be involved, *recA*-dependent and *recA*-independent (Bi & Liu, 1994; Lovett *et al.*, 1993). There seems to be a trend towards increasing *recA*-independence of deletion with decreasing repeat length and decreasing distance between the repeats. However, there may be replicon or sequence context effects that govern the contribution of a *recA*-independent mechanism for deletion.

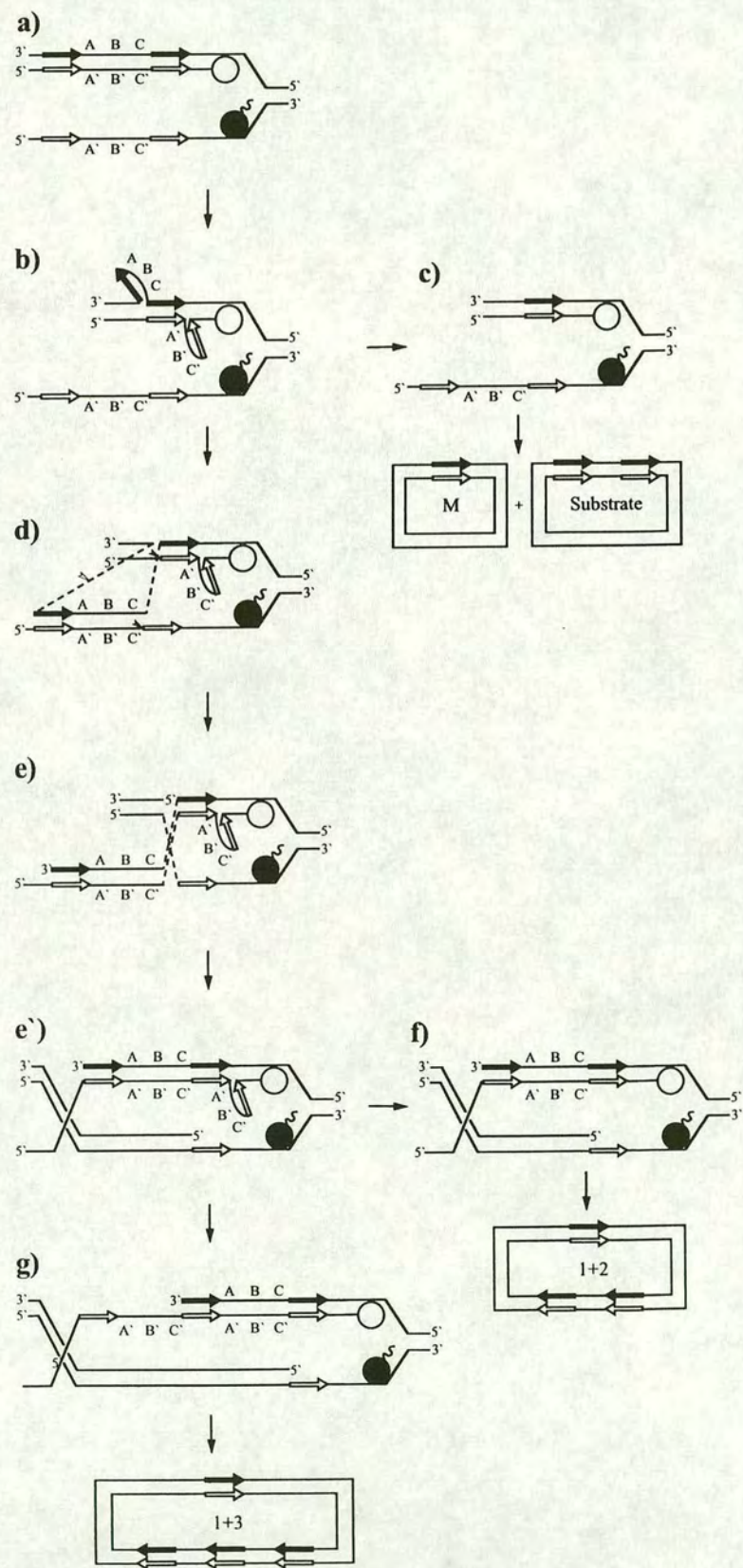
The predicted product of recombination between plasmid-borne direct repeats is a monomeric plasmid, which has deleted one repeat and any intervening DNA. However, many studies have found dimers to be the major product of *recA*-independent recombination (Dianov *et al.*, 1991; Mazin *et al.*, 1991; Lovett *et al.*, 1993; Bi & Liu, 1994; Bi *et al.*, 1995; Bi & Liu, 1996). Two hetero-dimeric forms are commonly observed, 1+2 and 1+3 (see Fig.1.3 for diagrammatic representation of these heterodimers). Bi & Liu (1996), demonstrated that the length of the repeat and the distance between the repeats affect the relative abundance of each form of product. They proposed a model for DNA recombination between direct repeats that combines both strand slippage during DNA replication, and *recA*-independent recombination (described in Fig.1.3). The model most similar to this one is the sister-strand-exchange model for the formation of dimeric products (Lovett *et al.*, 1993).

A review by Kuzminov (1996), highlighted a possible mechanistic link between deletion formation, associated with dimerization of plasmid DNA, and the mutation and evolution of animal mitochondrial DNA (mtDNA). This has an important medical aspect as many human degenerative diseases are associated with deficiencies in mitochondrial function, due to duplications (Poulton *et al.*, 1993) or deletions (Wallace, 1992) in mtDNA. Heterogeneous mtDNA deletions have been found to accumulate during the human ageing process, and are thought to arise from either strand slippage or another intramolecular illegitimate recombination event during replication (Baumer *et al.*, 1994).

Figure 1.3

A replication model for DNA recombination between direct repeats (Bi & Liu, 1996). Each strand of a repeat participating in recombination is shown as an arrow. (Note that the repeat shown may be only a segment of the entire duplication in the substrate). Open arrows correspond to the 5' to 3' orientation of the DNA and filled arrows to the 3' to 5' orientation. A, B & C are arbitrary markers. The open and shaded circles represent the leading and lagging strand polymerases, respectively. The wavy line represents the RNA primer for lagging strand DNA polymerase. An explanation of the model is as follows: a) At the replication fork and after replication of the repeats in the leading strands is completed, the region containing the direct repeats in the lagging strand remains single-stranded. b) If stalling of the replication fork occurs, misalignment may take place between the repeats in the nascent DNA helix of the leading strand and its template; this would generate two single-stranded loops, each consisting of one strand of a repeat and any intervening sequence. c) The loops may subsequently be removed by endonuclease and ligase; completion of replication will then lead to the formation of a monomeric product of deletion (M) together with a substrate plasmid. d) If the loop originated from the leading strand template anneals with its complementary sequence in the lagging strand template, dimeric products will be formed as discussed below. The open arrowhead indicates the specific junction created in the leading strand template that is to be cleaved. Independently, as indicated by the two filled arrowheads, the two specific junctions, one in the lagging strand template and the other in the leading strand are to be cleaved and later misjoined (i.e. rejoined after the ends are exchanged). Note that these processes do not have to occur simultaneously. Completion of these processes leads to the structure shown in e) which is better illustrated in e'). Up to this point, a formal crossover is achieved and there is only one repeat left in the original lagging strand and, therefore, both deletion and dimerization are achieved. f) If the remaining single-stranded loop as shown in e') is removed, the result of replication would generate the 1+2 dimeric product. g) If the loop persists and "slips" back to anneal with its complementary sequence, the 1+3 dimeric product would be formed.

Figure 1.3



Strand Slippage and Disease

Simple sequence repeats are ubiquitous and highly polymorphic in eukaryotic genomes. Recently, the molecular basis of ten human genetic disorders including fragile X syndrome, myotonic dystrophy, Kennedy's disease, Huntingtons disease, spinocerebellar ataxia type 1, and dentatorubral-pallidoluysian atrophy has been partially established. At the molecular level these genetic diseases are all characterised by the expansion of a simple trinucleotide repeat, CTG and CGG. As discussed previously, sequences which have the potential to form unusual secondary structure can stimulate replication slippage. It has been proposed that the expansion and contraction of di-, tri-, and tetranucleotide repeats could be facilitated by secondary structure formation. Therefore, the underlying mechanism for the associated disease could be replication slippage (reviewed by Richards & Sutherland, 1994; Sinden & Wells, 1992; Wells, 1996). In the review by Sinden and Wells (1992), a hypothetical mechanism for the expansion of triplet repeats was put forward (see Fig. 1.4). Trinucleotide repeats can form hairpin loops, triplex or K^+ -dependent quadruplex DNA, when single stranded (Usdin, 1995; Wells, 1996). When double stranded, trinucleotide repeats have been shown to form what is best described as toroid DNA (Shimizu, 1996). All these secondary structures may cause stalling of DNA replication.

A recent review by Bouffler *et al.* (1993), highlights the role of DNA repeats and their associated secondary structures in genomic instability and neoplasia. Many studies have indicated, direct repeats play a major role in the deletion breakpoints in tumorigenesis. The relative precision of such deletions implies that some form of DNA secondary structure is formed between the flanking repeats. This secondary structure may be facilitating either replication slippage past a small DNA loop or sequence specific recombination, possibly unequal sister chromatid exchange between repeats (Canning & Dryja, 1989; Meuth, 1990; Streisinger *et al.*, 1960). Deletions and insertions in the p53 tumor suppressor gene are often associated with inverted or direct repeats and runs of purines or pyrimidines (Greenblatt *et al.*, 1996).

As previously discussed many human degenerative diseases are associated with deletion and/or duplication of mtDNA. Most mtDNA deletions in mitochondrial disease and ageing, are flanked by short (4-13bp) direct repeats (Wallace *et al.*, 1991). It is thought that the mechanism of deletion in mtDNA arises from either strand slippage or another intramolecular illegitimate recombination event during replication (Baumer *et al.*, 1994).

Figure 1.4

A hypothetical mechanism for the expansion of triplet repeats (Sinden and Wells, 1992). The formation of the expanded triplet repeats could occur by replication slippage where there is a strong block to DNA replication. This blockage, represented by the shaded box, could be a tightly bound protein, a region of triplex DNA, or a tetra-stranded region formed within the triplet repeat, preventing the progression of the polymerase (open circle). The long single $(\text{CGG})_n$ or $(\text{CAG})_n$ strand may be stabilised by the formation of a duplex (or other non-B-DNA structure).

Figure 1.4

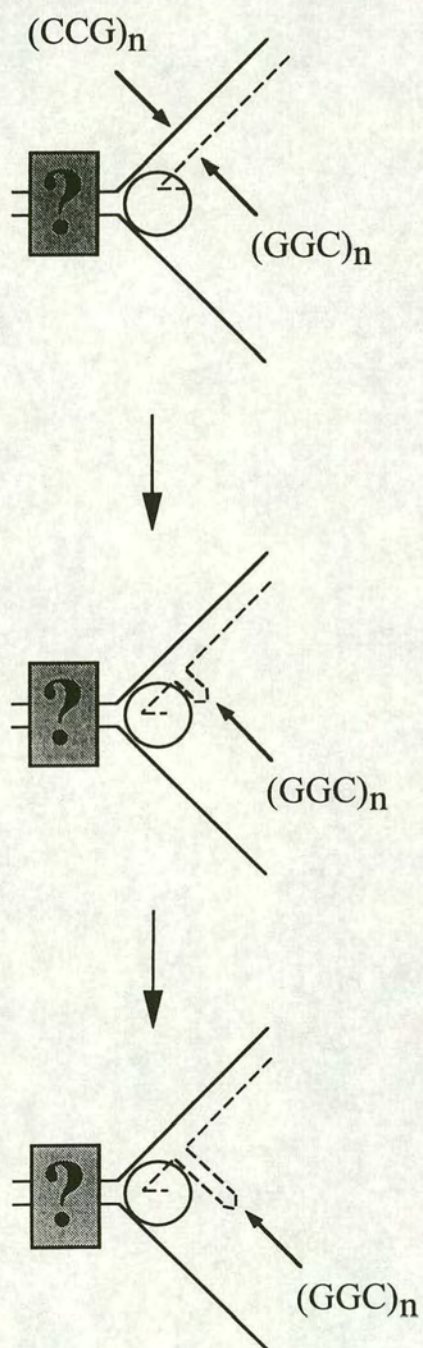
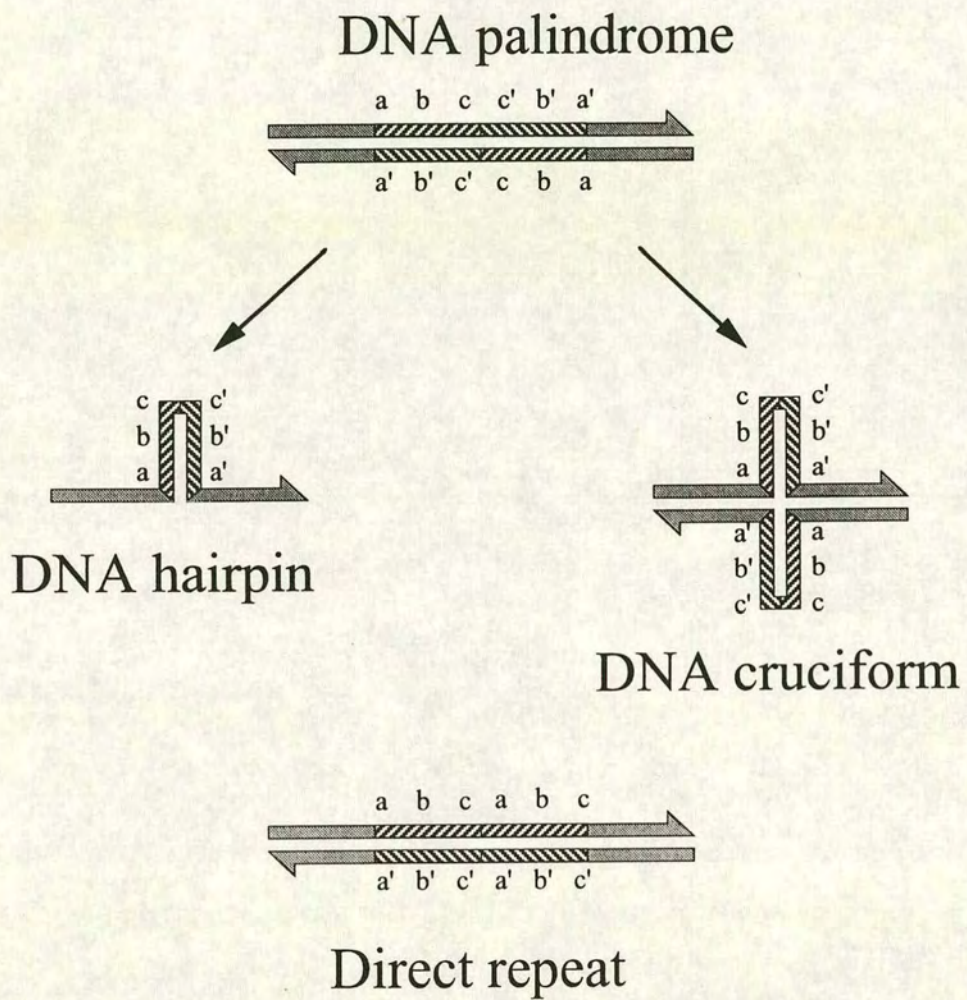


Figure 1.5

Potential DNA secondary structures formed by DNA palindromes (Leach, 1994).



DNA Palindromes

“DNA palindrome” is the term first used in 1974 by Wilson and Thomas, to describe inverted repeats with two-fold rotational symmetry. Palindromes have the potential to form secondary structure such as hairpins, when single stranded, or cruciforms, when double stranded (see Fig. 1.5). Long DNA palindromes are found frequently in eukaryotes where they can comprise up to several percent of the genome. In contrast, long DNA palindromes are not often found in prokaryotic genomes and are generally short and imperfect.

Long DNA palindromes have been shown to be unstable in eukaryotes (Collick, *et al.*, 1996; Gordenin, *et al.*, 1993; Henderson & Petes, 1993; Ruskin & Fink, 1993), prokaryotes (Peeters, *et al.*, 1988; Warren & Green, 1985; Weston-Hafer & Berg, 1991), and halt the progress of the replication fork *in vitro*, (Bedinger *et al.*, 1989; LaDuca *et al.*, 1983; Weaver & DePamphilis, 1984).

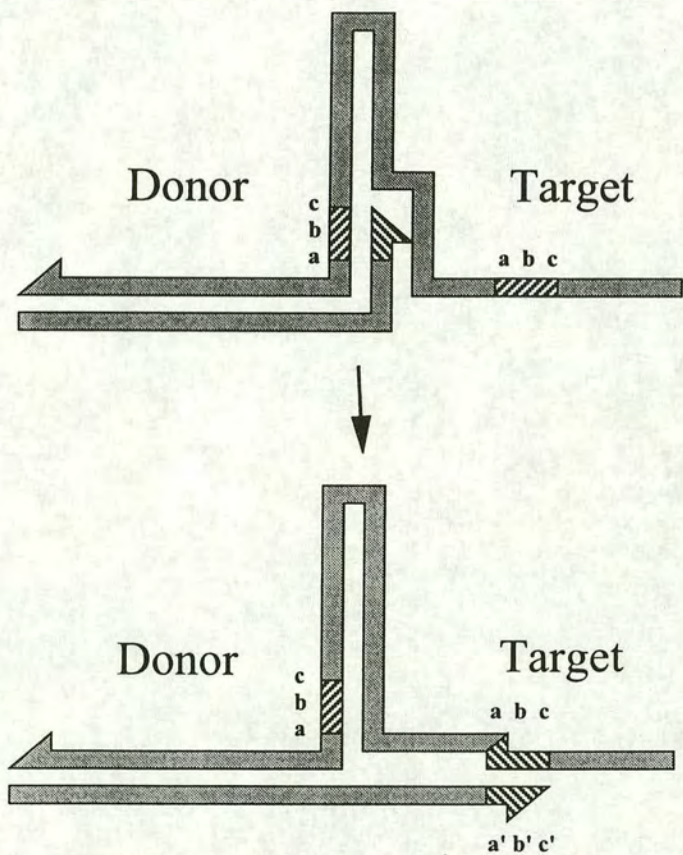
Possible Cellular Roles for Palindromic DNA

Evolution

Large genome rearrangements, such as duplications, deletions, translocations, inversions and insertions, are important for the evolution of a species. Palindromic DNA can mediate such rearrangements, an example of which, is its association with mammalian DNA amplification (Stark *et al.*, 1989). Palindromes have been shown to stimulate intrachromosomal recombination (reviewed by Klein, 1994). Palindrome-stimulated replication slippage, can facilitate deletions which can fuse different genes, creating proteins with novel regulatory control or function. For an overview of palindrome-stimulated replication slippage see Fig. 1.6.

Figure 1.6

Palindrome-stimulated replication-slippage. (Leach, 1994). When DNA replication enters a region of intra-strand base-pairing, there is an increased likelihood of stalling. The 3' end of the newly replicated strand can dissociate and pair downstream with another complementary sequence. This allows one to define a donor repeat and a target repeat in the strand slippage reaction.



Transcription & Replication

In a study by Park *et al.* (1993), the authors conclude that co-operative binding of two protein monomers, to its palindromic target site, increases the specificity of protein-DNA interaction. DNA palindromes are more common in regions that have regulatory function, than chance alone would predict (Muller & Fitch, 1982). Numerous studies (Watz & Pirrotta, 1975; L'Abbe *et al.*, 1991; Dimri & Das, 1990) have demonstrated the presence of palindromic DNA in transcriptional promoter regions. Palindromes have also been found associated with transcriptional terminators (reviewed by Platt 1986).

Palindromic DNA is a common feature of DNA replication origins. In *E. coli* the replication origin has both inverted and direct repeated sequence (Meijer *et al.*, 1979; Hirota *et al.*, 1979). Prokaryotic plasmids also have inverted repeat sequences at their replication origins (Lin & Meyer, 1987; Wang *et al.*, 1993).

Mobile Genetic Elements

Many mobile genetic elements (insertion sequences and many transposons) are flanked by almost perfect inverted repeats. The function of inverted repeats in mobile genetic elements is to allow the transposase to recognise the ends of the element to initiate strand-transfer (for a review see Berg & Howe, 1989).

Repetitive Palindromic DNA

In *E. coli*, *Salmonella typhimurium* and other members of the family *Enterobacteriaceae* repetitive palindromic DNA has been well characterised. Repetitive Extragenic Palindromes (REP sequences; Higgins *et al.*, 1982), have been identified as binding targets for a number of bacterial proteins (Yang & Ames, 1990; Gilson *et al.*, 1990), and are thought to have a primary role in chromosomal structure and organisation. Clusters of REPs are called Bacterial Interspersed Mosaic Elements (BIMEs; Gilson *et al.*, 1991). Repetitive IHF-binding Palindromes or Reiterative IHF BIME elements (RIP, Oppenheim *et al.*, 1993; or RIB, Boccard & Prentki, 1993), have a recognition site for a histone related protein and may insulate domains from

changes in supercoiling. Intergenic Repeat Units or Enterobacterial Repetitive Intergenic Consensus Sequences (IRUs or ERICS), have also been identified in *E. coli* (Sharples & Lloyd, 1990; Hulton *et al.*, 1991). Their function is at present undetermined, but it has been suggested that they represent a form of selfish DNA.

Human Genetic Disease

Trinucleotide repeat sequences have the potential to form hairpin-like structures. One hypothesis for the expansion of trinucleotide repeats, associated with genetic disease, is strand slippage. Stable secondary structure may facilitate strand-slippage by causing DNA replication to pause (for a review see Wells, 1996). In Prader-Willi syndrome long inverted repeats were shown to be associated with structural instability of chromosome 15 (Donlon *et al.*, 1986); the chromosomal lesions may have been due to cruciform extrusion. A recent study by Greenblatt *et al.* (1996), highlights the association of insertions and deletions in the p53 tumour suppressor gene in human cancers with, palindromes, tandem repeats and homocopolymer runs (runs of purines or pyrimidines). See Bouffler *et al.* (1993), for a review of the role of both inverted and direct repeats, and their associated secondary structures in genomic instability and neoplasia.

Long DNA Palindromes in *Escherichia coli*

Naturally occurring palindromes in *E. coli* are mostly short (<40bp) and generally imperfect. DNA palindromes longer than 150-200bp can not be cloned in wild type *E. coli*; either the replicon containing the palindrome is so poorly replicated that it is inviable (Warren & Green, 1985; Yoshimura *et al.*, 1986), or the palindrome is so unstable that it suffers partial or complete deletion (Collins, 1981; Hagen & Warren, 1983). These phenomena are interlinked and have been termed inviability and instability respectively. When cloning palindromes in *E. coli* various factors affect the levels of instability and inviability. These factors are reviewed in the following sections.

Palindrome Length

Generally an increase in palindrome length gives a corresponding increase in instability and inviability. When above the arbitrary limit of 150-200bp, palindromes cloned in *E. coli* tend to cause inviability. This observation may be due to a number of reasons. Perhaps, they extrude to form stable secondary structure, that either halts DNA replication completely. Or they may provide a suitable substrate for enzyme systems within the cell that remove secondary structure of this type. The SbcCD and RecBCD proteins may perform this function in *E. coli* (Leach, 1994). Palindromes are known to stimulate intrachromosomal recombination events (for a review see, Klein, 1994), therefore an inviable head-to-head dimer product similar to that observed by Mizuuchi (1982) may cause inviability (Bi & Liu, 1996).

Smaller palindromes (>200bp) tend to be viable in *Escherichia coli*, but may be unstable and prone to deletion. Palindromes as short as 22bp have been demonstrated to stimulate deletion (DasGupta *et al.*, 1987). Shafferman *et al.*, (1987), showed that a 68bp perfect palindrome is genetically unstable. Sequence analysis of the 21bp deletion derivatives revealed two deletion types. Their generation was best described by the formation of intermediate cruciform structures.

Sequence of Palindrome

The sequence of a DNA palindrome plays an important role in its genetic stability. The kinetics and energetics of cruciform extrusion (reviewed by Murchie & Lilley, 1992) are primarily determined by the sequence of the palindrome and the level of DNA supercoiling. DNA palindromes are believed to extrude to form cruciform structures by two pathways, S-type extrusion and C-type extrusion (see Allers 1993 and references therein for comprehensive review). S-type extrusion is characterised by a requirement for cations, relatively moderate activation entropies and enthalpies, and a profound effect of central sequence changes on the reaction kinetics. It involves the melting of a small (~10bp) region at the centre of a DNA palindrome in negatively supercoiled DNA. The single stranded region generated,

intrastrand base pairs to form a protocruciform structure. This transition state is stabilised by cation binding. Branch migration, driven by DNA supercoiling, leads to the formation of the full length cruciform. C-type extrusion is exhibited by a small group of DNA palindromes that extrude in low salt concentrations. In this extrusion type the absence of salt helps destabilise the DNA helix allowing the melting of large regions of DNA. The melted region containing the DNA palindrome can then intrastrand base pair to form a cruciform structure in one step. In order for palindromes to extrude by this pathway they, or the DNA to either side of the palindrome, must be very A/T-rich.

Asymmetry at the palindrome centre can alleviate inviability and instability. Collins *et al.* (1982), showed that the removal of the 2750bp central region of Tn5 to generate a perfect palindrome stimulates a five-fold increase in the precise excision of this transposon. A detailed study by Warren and Green (1985), demonstrated that insertions of approximately 50bp in the centre of a palindrome were required to overcome inviability, but insertions of at least 150bp were required to overcome instability.

Short Direct Repeats

Many studies of deletion stimulated by palindromic sequences have shown that their endpoints tend to occur in short direct repeats (Bichara, *et al.*, 1995; Collick, *et al.*, 1996; DasGupta, *et al.*, 1987; Peeters, *et al.*, 1988; Weston-Hafer & Berg, 1991). Most models for palindrome-stimulated deletion postulate that a cruciform or hairpin secondary structure formed by a palindrome brings direct repeats into close proximity, allowing deletion of the DNA between the repeats during DNA replication (Glickman & Ripley, 1984; Trinh & Sinden, 1993). Direct repeats are deletion hotspots in their own right. In a recent study by Bi and Liu (1996), recombination between direct repeats was examined. The frequency of recombination was affected by both the length of the repeat, and the distance between the repeats.

Genotype of Host

The propagation of palindromic sequences is affected by the genotype of the host. Leach and Stahl (1983), demonstrated that a λ phage carrying a long DNA palindrome could plaque on *recBC sbcB* hosts. Chalker *et al.* (1988), subsequently proved that this property was not a function of the *recBC* and *sbcB* mutations themselves. Instead, the ability to plate long DNA palindromes was due to a mutation in *sbcC*. Sequence analysis of *sbcC* revealed that it is co-transcribed with *sbcD* (Naom *et al.*, 1989). *E. coli sbcCD* mutants are known to allow the replication of DNA containing long palindromes (Collins *et al.*, 1982; Chalker *et al.*, 1988; Gibson *et al.*, 1992).

Connelly & Leach (1996), have demonstrated that the *sbcC* and *sbcD* genes of *E. coli* encode a nuclease activity. One possible hypothesis for palindrome mediated inviability is that *SbcCD* recognises unusual secondary structure and removes them from the replicon.

DNA Replication

DNA replication is closely associated with palindrome-mediated inviability and instability. When a λ phage with an 8.4kb palindrome infects an *E. coli rec⁺ dnaB^{ts}* and DNA replication is prevented by growth at the restrictive temperature, there is no loss of the palindrome sequence or its carrier replicon. However, growth at 37°C revealed, phage that have undergone DNA replication formed deletion derivatives of the palindrome (Shurvinton *et al.*, 1987). In another study by Lindsey and Leach (1988), demethylation was used to determine the fate of palindrome-containing λ DNA in an *E. coli rec⁺ dam* host. The results observed, implied that inviability was a consequence of slow replication.

DNA palindromes are known to halt the progress of the replication fork *in vitro*, (Bedinger *et al.*, 1989; LaDuca *et al.*, 1983; Weaver & DePamphilis, 1984). If the replication block is stable it can render the replicon containing the DNA

palindrome inviable. Pausing of DNA replication could also facilitate strand slippage or template switching between direct repeats.

SECTION 2:

Cosmid Cloning

Introduction

The Human Genome Project is a massive genome sequencing initiative aimed at obtaining the nucleotide sequence of the entire genome of man and its physical mapping. The analysis of the human genome is critically dependent on the techniques of DNA cloning in bacteria and yeast. A number of useful cloning vectors have been developed. These include, cosmids (Collins & Hohn, 1978), Yeast Artificial Chromosomes (YACs; Burke *et al.*, 1987), Bacterial Artificial Chromosomes (BACs; Shizuya *et al.*, 1992) and P1-derived Artificial Chromosomes (PACs; Ioannou *et al.*, 1994). At the present time, cosmids and YACs are the most practical vehicles for the cloning and manipulation of large DNA fragments.

Various sequences in mammalian genomes are unstable and subject to rearrangement when cloned. Many sequences are unstable because they have the potential to form unusual secondary structure (for an overview of unstable sequences, capable of forming secondary structure, see Section 1: Strand Slippage and Secondary Structure). The following section reviews cosmids, and to a lesser extent YACs, highlighting their uses, problems associated with the vectors, and developments.

Cosmid Vectors

Cosmid vectors were first developed by Collins and Hohn (1978), to overcome the problems associated with transforming *Escherichia coli* with large plasmids. Efficient transfer of DNA plasmids of up to about 15kb into competent *E. coli* is possible. However, molecules above this limit give a dramatic reduction of transformation efficiencies (Kushner, 1978). Cosmids may be defined as plasmids containing: a) the *cos* region of bacteriophage λ , required for *in vitro* packaging (Collins & Hohn, 1978; Hohn & Collins, 1980), b) an antibiotic resistance gene for selection, and c) unique restriction sites for insertion the foreign DNA.

Cosmids have been applied in many different procedures, including: the construction of high quality, representative genomic libraries (Little, 1989; Vincze & Kiss, 1990), restriction mapping (Evans & Wahl, 1987), physical mapping by chromosome walking or jumping (Bender *et al.*, 1983; Breter *et al.*, 1997; Spoerel & Kafatos, 1987) or sequence fingerprinting (Stallings *et al.*, 1990), expression of foreign genes in *E. coli* (Choo *et al.*, 1986) and P-element-mediated transformation in *Drosophila* (Speck *et al.*, 1988).

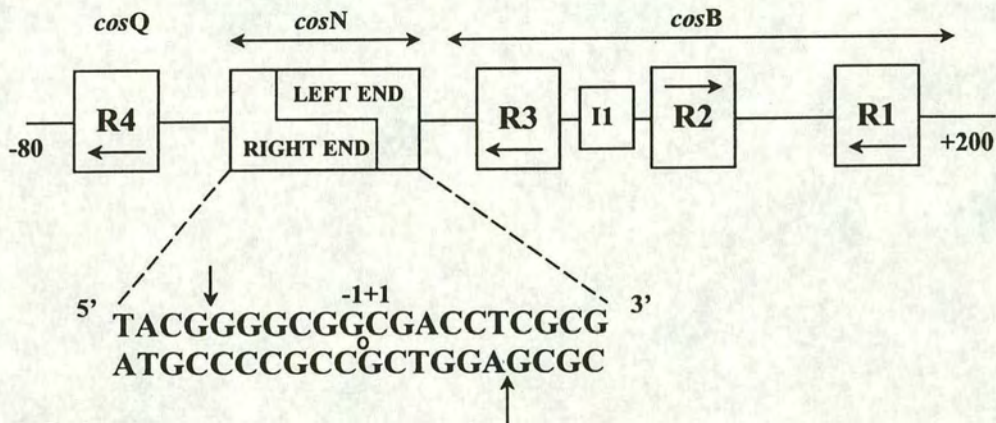
***In vitro* Packaging of Cosmid DNA**

An *in vitro* packaging assay for the study of “DNA State” when packaged, was developed by Hohn & Hohn (1974). It was demonstrated that only a small region of λ DNA (the *cos* site) in the proximity of the cleavage site is required for recognition by the packaging mechanism (Hohn, 1975). DNA fragments of approximately 37kb to 52kb are packaged into λ phage heads (Feiss *et al.*, 1977). These limits govern the cloning capacity of an individual cosmid. The cloning of DNA fragments in the range of 30-50kb are typical.

The cohesive-end site (*cos*) is where DNA packaging is initiated, proceeding in a vectoral manner from left to right along the chromosome. It is also where endonucleolytic cleavage of DNA takes place, separating concatemers into single chromosomes which are packaged into λ proheads (for a review see Murialdo, 1991). As summarised in Fig. 1.7, *cos* comprises about 200bp and consists of several subunits. They include: a) *cosB*, which contains a series of protein binding sites b) *cosN*, where the DNA is nicked specifically by terminase to generate the 12bp cohesive ends, and c) *cosQ*, which contains the R4 element, (originally classed as part of *cosB*). Terminase, the DNA packaging enzyme of bacteriophage λ is an ATP-stimulated site specific endonuclease, comprising the products of λ genes *NuI* and *A* (Higgins & Becker, 1995). Terminase binds to *cosQ* and *cosB*, to the left and right of *cosN* forming a DNA-protein complex called the termisome. The termisome complex recognises *cosN* and cuts the DNA to allow fragments of the correct size to

Figure 1.7

Summary of the *cos* cohesive-end site of bacteriophage lambda. Terminase nicks the DNA at *cosN* to regenerate the 12-base cohesive ends of the mature λ chromosome. The alternating orientation of the four gpNu1-binding sequences, called *R*-1,2,3+4, and one IHF binding sequence called *I1*, are also shown.



be packaged. *In vitro* packaging kits are available commercially. Using these kits it is possible to package cosmid DNA and subsequently infect and transform *E. coli* strains by manipulating the natural life cycle of lambda.

The Development of Cosmids

The first cosmid vectors developed were based on ColE1 hybrid plasmids carrying a cohesive-end site (Collins *et al.*, 1976). Many subsequent cosmids were based on the plasmid vector pBR322 (for a review see Balbás *et al.*, 1990). Cosmids have been constructed with a variety of structural elements designed to improve sequence representation, to simplify or accelerate the structural or functional analysis of cloned DNA, or to simplify the construction of high quality, representative genomic libraries. The features have included: selectable genes for transfection into and rescue from eukaryotic cells (Bates & Swift, 1983; Breter *et al.*, 1997; Choo *et al.*, 1986; Evans *et al.*, 1989; Grosveld *et al.*, 1992), features for rapid restriction mapping using phage λ terminase (Little & Cross, 1995; Rackwitz *et al.*, 1985), recognition sequences for specific oligonucleotides (Evans & Wahl, 1987; Ji *et al.*, 1996; Wahl *et al.*, 1987), or triple-helix-mediated affinity capture (Ji *et al.*, 1994; 1996), intron-encoded endonuclease sites (Ji *et al.*, 1996), rare cutting restriction endonuclease sites (Campbell, 1989; De Smet *et al.*, 1993; Ji *et al.*, 1996; Wahl *et al.*, 1987), altered cloning capacity (Saito & Stark, 1986), bacteriophage promoters for the synthesis of end-specific RNA probes for chromosome walking (Cross & Little, 1986; Evans & Wahl, 1987; Wahl *et al.*, 1987), bacterial insertion sequences and double origin vectors for simple preparation of DNA templates for sequencing (Ahmed, 1987; 1989; 1994), transcriptional terminators to prevent expression of functional genes when cloned (Gibson *et al.*, 1987), multi-copy replication origins (Ehrich *et al.*, 1987), features to simplify library construction by including two (Bates & Swift, 1983; Ishiura *et al.*, 1989; Ji *et al.*, 1996) or more (Pirrota *et al.*, 1983) *cos* sequences in the vector, specific sequences facilitating the removal of vector sequence by triple-helix-mediated affinity capture (De Smet *et al.*, 1993), and origins of replication that improve the stability of cloned sequences (Kim *et al.*,

1992; Jiang *et al.*, 1987; Little & Cross, 1985; Nakano *et al.*, 1995; Speek *et al.*, 1988) or enable the cloning of genes which are toxic to the host, in high copy number vectors (Jiang *et al.*, 1987; Nakano *et al.*, 1995). It is interesting to note that the cosmid vectors developed which seemed to help stabilise cloned sequences had either, low copy number origins of replication or origins which gave a constant copy number.

Cosmids have also been developed for use in a broad range of bacterial species. These include vectors specifically designed for use in: *Salmonella typhimurium* (Palva *et al.*, 1981), *Pseudomonas aeruginosa*, *P. putida*, *Alcaligenes*, *Thiobacillus*, *Rhizobium*, *Agrobacterium*, *Rhodopseudomonas*, *Vibrio cholerae* (Bagdasarian *et al.*, 1981; Frey *et al.*, 1983; Gallie *et al.*, 1985; Connell *et al.*, 1995), and *Mycobacterium smegmatis* (Hindshelwood & Stoker, 1992). Shuttle cosmid vectors have been developed for use in *E. coli* and eukaryotes. These include *Saccharomyces cerevisiae* (Breter *et al.*, 1987) and *actinomycetes* (Denis & Brzezinski, 1992). Recently a 'phasmid', which functions as a cosmid in *Escherichia coli*, and as a phage in *Streptomyces* has also been characterised (Zhou *et al.*, 1994).

Problems Associated with the use of Cosmids

In early work involving the first use of cosmids, Collins and Hohn (1978), observed the absence of palindromic structures; they also noted a high number of deletions among their isolates (4 out of 52). There have since been many anecdotal reports of instability of human DNA inserts in genomic cosmid libraries (Chia *et al.*, 1982; Steinmetz *et al.*, 1982; Yelton *et al.*, 1985).

Due to the presence of tandem repeats and other unstable sequences, human DNA often undergoes rearrangement when cloned in *E. coli*. Rearrangements due to homologous recombination can be prevented with the use of *recA* strains (Kurnit, 1989). Ishiura *et al.* (1989), demonstrated a *recB recC sbcB recJ* host prevented cosmid deletion events that occurred in *recA* hosts. However, they since have carried out a comprehensive study of RecA-independent deletion of recombinant cosmid

DNA in *E. coli* (Ishiura *et al.*, 1990), and concluded a *recB recC sbcB recJ* host did not prevent all deletion events. Ishiura *et al.* (1990), observed various sequence motifs close to the junctions of the deletion events investigated. These included: direct repeats, inverted repeats, Chi-related sequences and frequent 5-10bp stretches of each nucleotide. These sequences are all known to have an association with possible illegitimate recombination events; for a review see Section 1 of this thesis.

It has been suggested by a number of groups that low copy number vectors may improve the stability of cloned eukaryotic sequences (Kim *et al.*, 1992; Jiang *et al.*, 1987; Little & Cross, 1985; Nakano *et al.*, 1995; Speek *et al.*, 1988). It is possible that improved stability in low copy number vectors is partially true; as deletion events probably reflect two processes, the formation of deletions, and the displacement of original cosmids by deleted forms. It is likely that low copy number vectors appear to possess improved stability, because they generate more uniform genome representation. Some genes when cloned in high copy number vectors are toxic to *E. coli*. Low copy number vectors are also useful for the cloning of such sequences (Nakano *et al.*, 1995). Most groups conclude a low or intermediate copy number origin of replication to be the best choice when constructing a cosmid.

Fernandez *et al.* (1986), showed that spontaneous insertion of *E. coli* insertion element *IS1*, occurred during propagation of two vectors pNN1 and pTCF. He concluded that the use of a vector carrying *IS1* for cosmid cloning, caused a very high frequency of deletion in standard *recA* hosts, (this observation has been applied by Ahmed (1987; 1989), where production of deletions using *IS1*, allowed the sequencing of long DNA fragments in cosmids).

Cosmids are stability maintained and do not undergo the same levels of rearrangements in yeast (Hohn & Hinnen, 1980), consequently shuttle vectors for *E. coli* and *Saccharomyces cerevisiae* have been constructed (Breter *et al.*, 1987).

Yeast Artificial Chromosomes

YACs were first developed for use as a cloning vector by Burke *et al.* (1987). In general YACs consist of large (up to 1 megabase) inserts of DNA ligated between vector arms. The arms end in telomere sequences and contain centromeres (*CEN*), replication origins (*ARS*) and selectable markers (eg. *TRP1* or *URA3*) to stabilise the YACs in yeast (Burke *et al.*, 1987; Schlessinger, 1990). Many YACs have bacterial replication origins and selectable markers, to facilitate mapping and recovery of part of an insert as a plasmid in *E. coli*. Developments have included: the addition of T3 and T7 promoters, or rare restriction endonuclease sites, selectable markers for mammalian cells and YACs constructed for use in many bacterial and plant species, (Schlessinger, 1990; and references within). YACs are mainly used as tools for mapping and analysis of genomes in a similar fashion to cosmids.

Problems Associated with the use of YACs

Burke *et al.* (1987), suggested it was conceivable that sequences unstable when cloned in *E. coli*, may have improved stability in yeast. However, many studies have reported instability of repeated, and other sequences in plasmids (Bijovet *et al.*, 1991; Gordenin *et al.*, 1993; Henderson & Petes, 1993; Ruskin & Fink, 1993), and YACs (Matsuoka *et al.*, 1991; Neil *et al.*, 1990; Soeda, 1993). Other problems associated with YACs include: low transformation efficiency, the abundant presence of chimaeric clones in nearly all YAC libraries, and difficulties in DNA manipulation relative to bacterial systems (Baldini *et al.*, 1992; Bates *et al.*, 1992; Green *et al.*, 1991).

Summary

In most studies of genomic mapping and analysis, cosmids and YACs are used together. Both systems are subject to similar problems of instability due to illegitimate recombination and other events. In order to achieve the goals of the

Human Genome Project, there is a necessity for the development of a cloning system which challenges the effects of illegitimate recombination. At the moment many eukaryotic sequences are considered so unstable in *E. coli* and *Saccharomyces cerevisiae* as to be un-clonable. It is interesting to note a recent development, the possible construction of Mammalian Artificial Chromosomes (MACs). MACs would in theory be similar to YACs, containing two vector arms which end in human telomeres and markers for selection, replication origins and centromeres. However, replication origins and centromeres that are active in mammalian cells remain to be defined.

CHAPTER 2:

MATERIALS & METHODS

MATERIALS

Microbiological Strains, Media and Solutions

Strains

All *Escherichia coli* strains and plasmid vectors used in this work are described in Tables 2.1 and 2.2 respectively.

Media

Bacteriological Media

The following quantities are for 1 litre final volumes made up in distilled water and sterilised by autoclaving for 20 minutes at 15lb in⁻² unless otherwise stated.

L Agar, 10g Bacto-typtone (Difco), 5g yeast extract (Difco), 10g NaCl, 15g Bactro-agar (Difco), adjusted to pH 7.2 with NaOH.

L Broth, 10g Bacto-typtone (Difco), 5g yeast extract (Difco), 10g NaCl, adjusted to pH 7.2 with NaOH.

SOB, 20g Bacto-typtone (Difco), 5g yeast extract (Difco), 0.5g NaCl, 10ml 250mM KCl adjusted to pH 7.0 with NaOH.

SOC, SOB supplemented just before use (per 100ml), 1ml 2M glucose, 1ml 1M MgCl₂ and 1ml 1M MgSO₄.

Media Additives

1M KCl Stock, made up in distilled water, autoclaved.

2M Glucose Stock, made up in distilled water, filter sterilised.

1M MgCl₂ Stock, made up in distilled water, autoclaved.

1M MgSO₄ Stock, made up in distilled water, autoclaved.

Ampicillin (100 mg ml⁻¹) Stock, (Beecham Pharmaceuticals) was stored at -20°C.

Kanamycin sulphate (50 mg ml⁻¹) Stock, (Beecham Pharmaceuticals) was stored at -20°C.

Xgal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) Stock, (Sigma) was made up in dimethylformamide and stored at -20°C. It was used at 40 μg ml⁻¹.

IPTG (Isopropylthio-β-D-galactoside) Stock, (Sigma) 1M, was made up in sterile distilled water and stored at -20°C.

Materials for DNA Purification and Manipulation

Unless otherwise stated, general laboratory chemicals were purchased from Sigma, Fisons or BDH.

General Solutions and Materials for DNA Purification and Manipulation

1M Tris/HCl Stock (pH 7.5), 1M Tris base, adjusted to pH 7.5 with concentrated HCl, autoclaved.

0.5 M EDTA Stock (pH 7.5), 0.5 M EDTA disodium salt, adjusted to pH 8.0 with glacial acetic acid, autoclaved.

TE Buffer Stock (10x), 100mM Tris, 10mM EDTA, adjusted to pH 7.5 with concentrated HCl, autoclaved.

Solutions for Ethanol Precipitation

3M Sodium Acetate (pH 7.0), 3M sodium acetate, adjusted to pH 7.0 with glacial acetic acid, autoclaved.

70% Ethanol, prepared by adding 0.3 volumes of sterile distilled water to 0.7 volumes of ethanol.

Materials for Purification of DNA from Agarose Gels and Solutions

DNA in agarose gels or solutions were purified using either GENECLAN[®] II kit (Bio.101) or QIAquick[™] Gel Extraction Kit (QIAGEN 28704). The solutions used were those supplied by the kit.

Solutions for Plasmid/Cosmid DNA Purification

Rapid Plasmid Preparation

Li Buffer, 2.5M LiCl, 50mM Tris-HCl (pH 8.0), 4% (v/v) Triton X-100, 62.5mM Na₂EDTA, filter sterilised.

QIAGEN Plasmid Purification

Large scale plasmid DNA preparation was performed using a QIAGEN Plasmid Midi Kit. The solutions used were those supplied by the kit.

SDS Alkaline Lysis Plasmid DNA Purification

TEG, 0.9% glucose, 25mM Tris-HCl (pH 7.5), 10mM EDTA, filter sterilised.

Alkaline SDS, 0.2M NaOH, 1% SDS (sodium dodecyl sulphate), freshly prepared.

Potassium Acetate, potassium acetate (3M in potassium, 5M in acetate) was prepared by adding 0.4 volumes of sterile 5M acetic acid to 0.6 volumes of sterile 5M potassium acetate, autoclaved.

TE-RNAase A, 10mM Tris/HCl (pH 7.5), 1mM EDTA, 20µg ml⁻¹ RNAse A. prepared by adding 100µl RNAse A stock (10mg ml⁻¹) to 50ml sterile 1xTE buffer. Stored at 4°C.

Triton Lysis Midi Plasmid DNA Purification

10% Triton X-100, 10% (v/v) Triton X100 was made up in sterile distilled water.

Phenol-Chloroform, Phenol-Chloroform was prepared by adding an equal volume of 24:1 Chloroform-Isoamylalcohol (see below) to equilibrated phenol (phenol saturated with 10xTE buffer).

Chloroform-Isoamylalcohol, 24 volumes of chloroform were added to 1 volume of isoamylalcohol, stored in dark bottle to protect from light.

Preparation of Cosmid DNA from Minilysates

STET Buffer, 8% sucrose, 5% Triton X-100, 50mM EDTA, 50mM Tris-HCl (pH 8.0), in sterile distilled water.

Enzymes, Buffers and Solutions for DNA Manipulation

Restriction Enzymes

All restriction endonucleases used in this work and their reaction buffers were purchased from NBL Gene Sciences Limited, New England Biolabs, Boehringer Mannheim or Promega, and used according to the manufacturers instructions.

DNA Sequencing

DNA was sequenced using a Sequenase[®]v2.0 Sequencing Kit (United States Biochemical). The solutions used were those provided by the kit. Sequencing primers were manufactured by Oswel DNA service, or Perkin Elmer Econopure[™].

Other Enzymes

Reaction buffers were supplied by the manufacturer.

Bacteriophage T4 DNA ligase, (New England Biolabs) was incubated in 50mM Tris-HCl (pH 7.5), 10mM MgCl₂, 10mM DTT, 1mM ATP, 25µg ml⁻¹ BSA.

Shrimp Alkaline Phosphatase (SAP), (United States Biochemical) was incubated in restriction endonuclease buffer, or 200mM Tris-HCl (pH 8.0), 100mM MgCl₂.

Klenow Enzyme, (Boeringer Mannheim) was incubated in restriction endonuclease buffer.

Other Solutions

dNTP Stocks, (Boeringer Mannheim) were prepared in sterile distilled water at concentration of 50mM and 2mM. Stored at -20°C.

BSA, (20mg ml⁻¹) bovine serum albumin (NBL Gene Sciences Limited) was stored at -20°C.

Solutions for Gel Electrophoresis

Agarose Gel Electrophoresis

TAE Gel Buffer Stock, (20x), 0.8M Tris/acetate, 20mM EDTA (pH 8.0).

TAE Gel-Loading Buffer Stock, (5x), 0.2M Tris/acetate, 0.25M EDTA (pH 8.0), 0.2% bromophenol blue, 15% Ficoll 400.

Ethidium Bromide, (10mg ml⁻¹), prepared in distilled water and used at 0.5µg ml⁻¹.

Polyacrylamide Gel Electrophoresis

TBE Gel Buffer Stock, (10x) 0.89M Tris-borate, 20mM EDTA (pH 8.0).

Formamide-EDTA Gel Loading Buffer, 98% formamide, 0.1% bromophenol blue, 0.1% xylene cyanol FF, 10mM EDTA (pH 8.0).

Stop Solution, 95% formamide, 10mM NaOH, 0.25% bromophenol blue, 0.25% xylene cyanol FF.

Longranger™ Gel Solution, 50% stock (AT Biochem).

MDE™ Gel Solution, 2X stock (AT Biochem).

TEMED, (N-N-N'-N'-tetra-methyl-1,2-diamino-ethane)(Sigma Chemical Company) was stored at 4°C and protected from light.

10% AMPS, (ammonium persulphate) was freshly prepared in sterile distilled water.

Solutions for Silver Staining

The solutions used for silver staining were those included in the Silver Stain Plus Kit (BioRad).

Solutions for Cosmid Methods

***In Vitro* Packaging of Cosmid DNA**

Cosmids were *in Vitro* packaged by using the Gigapack[®] II XL Packaging Extract kit (Stratagene). The solutions used were those supplied by the kit.

SM Buffer, 0.1M NaCl, 0.01M MgSO₄·H₂O, 0.05M Tris-HCl (pH 7.5), 0.01% (w/v) gelatin.

Human Genomic DNA, CLONTECH Laboratories, Inc., source: placental tissue.

Transformation

CaCl₂, (0.1M), made up in distilled water, autoclaved.

Electroporation

HEPES, (1mM), (N-2-hydroxyethylpiperazine-N'-2ethanesulfonic acid) (pH 7.5) prepared in sterile MilliQ water.

Table 2.1

Escherichia coli Strains Used in this Work.

Strain	Description	Reference/Source	Notes
JM83	<i>ara</i> Δ (<i>lac-proAB</i>) <i>rpsL</i> (Str ^r) Φ 80 <i>lacZ</i> Δ M15	Yanish-Perron <i>et al.</i> , 1985	
DL494	<i>sbcC201 phoR::Tn10</i>	D. Leach	1
DL733	Δ <i>sbcCD::Kan^r</i>	D. Leach	1
DL795	<i>sbcC201</i> Δ (<i>mcrBC mrr hsdRMS</i>) <i>mcrA</i> <i>recA::Cm^r</i>	D. Leach	2
1046	<i>recA</i>	Bates & Swift, 1983	
DL901	<i>sbcC201 phoR::Tn10</i> Δ <i>pcnB::Kan^r</i>	D. Leach	1
DL887	<i>recA::Cm^r</i>	D. Leach	1
GM119	F ⁻ <i>dam-3 dcm-6 metB1 galK2 galT22 lacY1</i> <i>tsx-78 supE44 thi-1 tonA31 mtl-1</i>	D. Leach	

1. Derivative of JM83, (Yanish-Perron *et al.*, 1985).
2. This strain is a derivative of K12SH28, which carries a mutation in the *deoA* gene (also known as *tpp*) encoding thymidine phosphorylase activity (Fangman & Novick, 1966).

Table 2.2 Plasmid/Cosmid Vectors Used in this Work

Plasmid	Markers	Reference/Source	Notes
pUC4K	Amp ^r Kan ^r	Vieira & Messing 1982	
pUC18	Amp ^r	Vieira & Messing, 1982	
pAC2	Amp ^r	A. Chalker	1
pMS5	Amp ^r	M. Shaw	2
pMS7	Amp ^r	M. Shaw	2
pM*	Amp ^r	This work	3
pM*167	Amp ^r	This work	4
pM*170	Amp ^r	This work	4
pM*171	Amp ^r	This work	4
pM*180	Amp ^r	This work	4
c2XB	Amp ^r Kan ^r	Bates & Swift, 1983	
cDRI	Amp ^r Kan ^r	This work	5
cDRII	Amp ^r Kan ^r	This work	6
cDRIII	Amp ^r Kan ^r	This work	6

1. Derivative of pUC18 (Vieira & Messing, 1982); clones have highly sectorized phenotype on AXI media.
2. Derivative of pUC18 (Vieira & Messing, 1982) originally thought to contain a 'mutation' which stabilised a 571bp DNA palindrome (C. Blake, M. Shaw & D. Leach, unpublished observations).
3. Derivative of pMS7, has 571bp palindrome removed.
4. Derivative of pM*, have different palindromes cloned into the *EcoRI* site.
5. Cosmid cloning vector derived from pM* and c2XB (Bates & Swift, 1983).
6. Cosmid cloning vector derived from pM*.

METHODS

DNA Purification and Manipulation

Ethanol Precipitation

Ethanol precipitation was used to concentrate and purify DNA. To 0.9 volumes of DNA solution, 0.1 volumes of 3M sodium acetate pH 7.0 and 2 volumes of freezer-cold (-20°C) ethanol were added. The contents of the tube were mixed by inversion and incubated at -20°C for at least an hour to allow precipitation of the DNA. This was followed by centrifugation to pellet the DNA at 15krpm for 30 minutes at 4°C (Sorvall Microspin 24 centrifuge). The ethanol was removed and replaced by a large volume of freezer-cold 70% ethanol. The mixture was then centrifuged at 15krpm for 10 minutes at 4°C (Sorvall Microspin 24 centrifuge). Following the removal of the 70% ethanol the DNA was allowed to air dry on the bench next to a Bunsen flame for 10 minutes and redissolved in 1xTE buffer, 1xTris/HCl buffer or distilled water.

Purification of DNA in Agarose Gels or Solutions

DNA in agarose gels or solutions were purified using either GENECLAN[®] II kit (Bio 101) or QIAquick[™] Gel Extraction Kit (QIAGEN 28704).

GENECLAN[®] II Kit

This kit is based on the method of Vogelstein and Gillespie (1979). The DNA fragment of interest was excised from the agarose gel using a sterile scalpel blade. The band was weighed in a tar ed Eppendorf tube to give an estimate of volume. Into the Eppendorf tube 3 volumes of NaI stock solution were added and the agarose gel dissolved by incubating at 45° to 55°C for 5 minutes. Where DNA was in solution 3 volumes of NaI stock solution were added and the incubation step at 45° to 55°C was

omitted. 5µl of GLASSMILK suspension were added, mixed by vortexing and incubated on ice for 5 minutes. The GLASSMILK/DNA complex was pelleted by centrifugation for 1 minute at 15krpm (Sorvall Microspin 24 centrifuge) and the supernatant removed. The pellet was washed 3 times by resuspending it in 500µl Freezer-cold NEW WASH followed by centrifugation for 1 minute at 15krpm (Sorvall Microspin 24 centrifuge). After the third wash all the NEW WASH solution was removed and the pellet resuspended in 10µl of 1xTE or distilled water and incubated in a 50°C water bath for 2.5 minutes. It was then centrifuged for 1 minute at 15krpm (Sorvall Microspin 24 centrifuge) and the supernatant containing the DNA was transferred to a fresh sterile Eppendorf tube. This step was repeated and the supernatant combined. The DNA was stored at -20°C.

QIAquick™ Gel Extraction Kit

The DNA fragment of interest was excised from the agarose gel using a sterile scalpel blade. The band was weighed in a tared Eppendorf tube to give an estimate of volume. Into the Eppendorf tube 3 volumes of QX1 buffer were added (for > 2% gels 6 volumes of QX1 buffer were added) and the agarose gel dissolved by incubating at 45° to 55°C for 10 minutes. Where DNA was in solution, 3 volumes of QX1 buffer were added and the incubation step at 45° to 55°C was omitted. After solubilising the agarose gel, the pH of the solution was checked with a pH strip. If the pH was > 7.5 then 10µl of 3M sodium acetate, pH 5.0 were added and the solution mixed. One gel volume of isopropanol was added, mixed by vortexing and the sample loaded onto a QIAquick spin column in a 2ml collection tube followed by centrifugation for 1 minute at 15krpm (Sorvall Microspin 24 centrifuge). After centrifugation the flow-through was discarded, 0.5ml of buffer QX1 was loaded onto the QIAquick spin column followed by centrifugation for 1 minute at 15krpm (Sorvall Microspin 24 centrifuge). The DNA bound to the QIAquick membrane was washed by loading 0.75ml of buffer PE to the QIAquick spin column incubating at room temperature for 5mins followed by centrifugation for 1 minute at 15krpm (Sorvall Microspin 24 centrifuge). After centrifugation the flow-through was

discarded and the centrifugation step repeated in order to remove residual wash. The 2ml collection tube *were* discarded and the QIAquick spin column was placed in a fresh 1.5ml Eppendorf tube. Depending on the downstream application the DNA was eluted from the QIAquick column by loading 50-20µl 1xTE, 1xTris/HCl or distilled water followed by centrifugation for 1 minute at 15krpm (Sorvall Microspin 24 centrifuge). After the final centrifugation the DNA was placed in a fresh sterile Eppendorf tube and stored at -20°C.

Plasmid/Cosmid DNA Purification

Rapid Plasmid Preparation

This method is based on a method from He *et al.* (1989). 1.5ml of a fresh overnight culture were transferred to an Eppendorf tube and centrifuged in the cold room (4°C) at 15krpm for 2 minutes (Sorvall Microspin 24 centrifuge). The supernatant was removed completely and the pellet resuspended in 200µl of Li buffer, to which 200µl phenol/chloroform were added. After vortexing for 10-15 seconds the cell debris was removed by centrifugation at 15krpm for 3 minutes at 4°C (Sorvall Microspin 24 centrifuge). 150µl of supernatant containing the plasmid DNA was removed and to which 300µl of freezer-cold ethanol were added. After centrifugation at 15krpm for 15 minutes at 4°C (Sorvall Microspin 24 centrifuge), the supernatant was discarded and the DNA pellet was washed in 800µl of freezer-cold 70% ethanol, followed by a further centrifugation at 15krpm for 5 minutes at 4°C (Sorvall Microspin 24 centrifuge). The DNA pellet was air dried for 10mins, redissolved in 40µl 1xTE RNase A and stored at -20°C.

QIAGEN Plasmid Purification Method

Using a QIAGEN Plasmid Midi Kit this method was used when large amounts of plasmid DNA *were* required. An appropriate volume (depending on the vector type) of fresh overnight culture was centrifuged at 12krpm for 6 minutes at

4°C (Sorvall centrifuge, GSA or SS34 rotor). The bacterial pellet was resuspended in 4ml of buffer P1. 4ml of buffer p2 were added, mixed gently, and then incubated at room temperature for 5 minutes. 4ml of chilled (0°C) buffer P3 were added, mixed and then incubated on ice for 15 minutes. After incubation the samples were centrifuged at 16krpm for 40 minutes at 4°C (Sorvall centrifuge, GSA or SS34 rotor). Following equilibration of a QIAGEN-tip 100 with 4ml of QBT buffer, the supernatant was removed promptly and introduced to the QIAGEN-tip 100. After the QIAGEN-tip had emptied the QIAGEN-tip was washed by passing two 10ml volumes of QC buffer through. The DNA was eluted from the column with 5ml of buffer QF. The DNA in solution was precipitated with 0.7 volumes of isopropanol which had been previously equilibrated to room temperature, by centrifugation at 16krpm for 40 minutes at 4°C (Sorvall centrifuge, GSA or SS34 rotor). The supernatant was discarded and the pellet washed with 2ml of freezer-cold 70% ethanol. After further centrifugation at 16krpm for 10 minutes at 4°C, the 70% ethanol was carefully poured off and the DNA pellet air dried for 15 minutes. The plasmid DNA was redissolved in 100µl 1xTE buffer and stored at -20°C.

SDS Alkaline Lysis Plasmid DNA Purification

This method is derived from Birnboim and Doly (1979). 1.5ml of a fresh overnight culture were transferred to an Eppendorf tube and centrifuged for 30 seconds at 15krpm (Sorvall Microspin 24 centrifuge). The supernatant was discarded and the tube centrifuged again for 5 seconds. The remaining liquid was removed and the pellet resuspended in 100µl ice-cold TEG by vortexing for 15 seconds. The cell suspension was incubated at room temperature for 5 minutes. 200µl of freshly prepared alkaline SDS were added, the tube was mixed by inversion and incubated for 5 minutes on ice. 150µl ice-cold potassium acetate were added and the tube was vortexed for 5 seconds. After 5 minutes on ice, the cell debris was removed by centrifugation at 4°C for 5 minutes at 15krpm (Sorvall Microspin 24 centrifuge). 400µl of supernatant containing the plasmid DNA were transferred to a fresh Eppendorf tube and extracted with equal volumes of phenol-chloroform and

chloroform. Two volumes of room temperature 100% ethanol were added, the tube vortexed briefly and incubated at room temperature for 5 minutes. The DNA was pelleted by centrifugation at 15krpm for 5 minutes at 4°C (Sorvall Microspin 24 centrifuge). The supernatant was discarded and the DNA pellet was washed in 800µl of freezer-cold 70% ethanol, followed by a further centrifugation at 15krpm for 5 minutes at 4°C (Sorvall Microspin 24 centrifuge). The DNA pellet was air dried for 10mins, redissolved in 40µl 1xTE RNase A and stored at -20°C.

Triton Lysis Midi Plasmid DNA Purification

25ml of fresh overnight culture was centrifuged at 12krpm for 6 minutes at 4°C (Sorvall centrifuge, GSA or SS34 rotor). The supernatant was discarded and the bacterial pellet resuspended in 500µl ice-cold 50mM Tris-HCl which were transferred to a fresh Eppendorf tube. A further 500µl of ice-cold 50mM Tris-HCl were used to rinse out the centrifuge tube. 50µl of freshly made lysozyme were added, the tube was mixed by inversion and incubated for 5 minutes on ice. 100µl of 0.5M EDTA pH8.0 were added, the tube was mixed by inversion and incubated for 5 minutes on ice. 10µl RNase were added, the tube was mixed by inversion and incubated for 5 minutes on ice. 20µl 10% Triton X-100 were added, the tube was mixed by inverting the Eppendorf tube 3 times and the cell debris was removed by centrifugation at 4°C for 15 minutes at 15krpm (Sorvall Microspin 24 centrifuge). The supernatant containing the plasmid DNA was transferred to a fresh Eppendorf tube and extracted with equal volumes of phenol-chloroform and isoamylalcohol-chloroform. An equal volume of freezer-cold 100% ethanol and 1/10 volume of sodium acetate pH7.0 were added and the tube vortexed briefly. The DNA was pelleted by centrifugation at 15krpm for 20 minutes at 4°C (Sorvall Microspin 24 centrifuge). The supernatant was discarded and the DNA pellet was washed in 500µl of freezer-cold 70% ethanol, followed by a further centrifugation at 15krpm for 15 minutes at 4°C (Sorvall Microspin 24 centrifuge). The DNA pellet was air dried for 10mins, redissolved in 100µl 1xTE RNase A and stored at -20°C.

Preparation of Cosmid DNA from Minilysates

This method is derived from Evans and Wahl (1987). 2.5ml of LB broth containing 20µg/ml ampicillin were incubated at 37°C for 6-8 hours. 1.5ml of the culture were transferred to an Eppendorf tube and centrifuged at 15krpm for 2 minutes (Sorvall Microspin 24 centrifuge). The supernatant was removed completely and the pellet resuspended in 300µl STET buffer. 25µl of freshly made lysozyme solution were added. After vortexing, the Eppendorf tube was placed in boiling water for 2 minutes. The solution was allowed to cool for 2-5 minutes and was centrifuged at 15krpm for 10 minutes (Sorvall Microspin 24 centrifuge). The gelatinous pellet was removed using a sterile toothpick and 325µl of 2-propanol were added to precipitate the plasmid DNA. After incubation at room temperature for 5 minutes the DNA was pelleted by centrifugation at 15krpm for 10 minutes (Sorvall Microspin 24 centrifuge). After centrifugation all the 2-propanol was removed and the DNA pellet was air dried for 10mins, redissolved in 25µl 1xTE RNase A and stored at -20°C.

General DNA Manipulation

Annealing of Oligonucleotides

Oligonucleotides (Oswel DNA service) were annealed in 1xTE buffer that had been made 10mM NaCl. This was prepared by adding 40µg of the oligonucleotide to 8 µl of 10xTE buffer and 32µl of 25mM NaCl in an Eppendorf tube. The mixture was heated in boiling water for 1 minute and allowed to cool slowly to room temperature. It was then diluted 100-fold in 1xTE Buffer, 10mM NaCl and stored at -20°C.

Restriction Enzyme Digests

Restriction enzyme digests were carried out according to the manufacturer's instructions, in the appropriate 10x incubation buffer with 5-fold excess of restriction



enzyme. Incubations were normally for 1 hour at the optimum temperature (usually 37°C).

DNA Sequencing

DNA was sequenced using a Sequenase[®] v2.0 Sequencing Kit (United States Biochemical). This method is derived from Sanger *et al.* (1977). 2-5µg plasmid DNA were purified using GENECLAN[®] II kit (Bio 101) or QIAquick[™] Gel Extraction Kit (QIAGEN 28704). 1-3µl (~0.5µg) sequencing primer (Oswel DNA service or Perkin Elmer Econopure[™]) and 7-9µl purified template DNA were added to an Eppendorf tube. The DNA was denatured by incubation at 100°C for 4 minutes, and snap-cooling on dry ice. Into a fresh Eppendorf 2µl of DTT, 2µl of Sequenase[®] reaction buffer, 0.7µl 1x labelling mix, 0.5µl (5µCi) [α^{35} S]dATP (Amersham) and 2.5µl diluted 1/8-1/5x Sequenase[®] enzyme were added. This labelling reaction was kept on ice while the template/primer mix was thawed and centrifuged briefly at 15krpm. The template/primer mix was then added to the labelling reaction and incubated at 15°C for 4 minutes. 4µl aliquots of the reaction were then transferred to 4 Eppendorf tubes, each containing 2µl of either ddT, ddC, ddA or ddG termination Mix; the tubes had been prewarmed to greater than 37 and less than 50°C before the addition of 4µl reaction mix. The termination reactions were carried out at 42°C for less than 5 minutes and halted with the addition of 4µl stop solution. The sequencing reactions were stored at -20°C for no more than a week.

DNA Ligation

DNA ligation was carried out in a total volume of no bigger than 20µl bacteriophage T4 DNA ligase buffer. To the DNA 50 units of bacteriophage T4 ligase (New England Biolabs) were added and the mixture was incubated for 1 hour at 37°C or for greater than 4 hours at 16°C. The reaction was terminated by incubation at 70°C for 10 minutes. If the DNA was to be purified (GENECLAN[®] II

kit (Bio 101) or QIAquick™ Gel Extraction Kit (QIAGEN 28704) the incubation step at 70°C was omitted.

DNA Dephosphorylation

DNA dephosphorylation was carried out immediately after DNA restriction. As Shrimp alkaline phosphatase is active in virtually all restriction enzyme buffers, the incubation buffer was not changed. 1-2µl (4U/µl) of Shrimp Alkaline Phosphatase (United States Biochemical) were added to the DNA restriction reaction and incubated at 37°C for 1 hour. The reaction was terminated by incubation for 15 minutes at 65°C.

Radiolabelling of DNA

Radiolabelling of DNA was generally carried out after an *EcoRI* digest. As Klenow enzyme is active in virtually all restriction enzyme buffers, the incubation buffer was not changed. 1µl (10µCi) [α^{35} S]dATP (Amersham), 1µl dTTP (2mM) and 1µl Klenow were added to 20µl DNA. The mixture was incubated at room temperature for 15 minutes then 1µl Klenow and 0.5µl chase (12.5 mM of each dNTP) were added and the mixture incubated for a further 10 minutes. The radiolabelled DNA was purified using a QIAquick™ spin column and eluted in 1xTE buffer.

Gel Electrophoresis

Agarose Gel Electrophoresis

Agarose gels were made and run in 1xTAE or 1xTBE gel buffer. The concentration of agarose was varied according to the size of the relevant DNA fragment (0.3% - 4% agarose). After 0.2 volumes loading buffer were added the

samples were loaded in the gel slots. Gels were generally run at $\sim 5\text{--}8\text{ V cm}^{-1}$ and examined on a C-62 Blak-Ray transilluminator (Ultraviolet Products Incorporated) after electrophoresis and Staining with $0.5\mu\text{g ml}^{-1}$ ethidium bromide. Pictures were taken using GRAB-IT™ (UVP,INC.).

Polyacrylamide Gel Electrophoresis

Polyacrylamide gels (40cm x 21cm x 0.4mm) were made and run in TBE gel buffer using a Sequi-Gen® Nucleic Acid Sequencing Cell (BioRad). The apparatus was assembled according to the manufacturer's instructions.

Sequencing Gels

DNA Sequencing gels were made by adding 20g urea to 11.2ml 5x TBE, 4.7ml Long Ranger™ (AT Biochem) and 16.4ml distilled water. The mixture was incubated at 50°C until all the Urea had dissolved. To 10ml of the mixture were added 50 μl TEMED and 125 μl freshly prepared 10% AMPS. This was used to seal the bottom of the Sequencing Cell. To the remaining volume of acrylamide were added 36 μl TEMED and 100 μl freshly prepared 10% AMPS. The acrylamide solution was poured into the Sequencing Cell using a 25ml glass pipette. The flush side of a 24-tooth shark's-tooth comb was inserted, the top of the cell covered with Saran wrap and the Gel left to polymerise for at least 1 hour. The Gel was assembled in the Sequencing Cell apparatus and pre-run to a temperature of $\sim 50^{\circ}\text{C}$ at 40-50W for 1 hour. The shark's-tooth comb was removed and reinserted toothed side first. The DNA sequencing reactions were denatured by incubation at 100°C for 3 minutes and snap-cooled on ice. 5 μl of each reaction were loaded onto the gel in the order: ddT, ddC, ddA, ddG. The gel was run at a constant power of 40W for 1 hour to 3 hours according to the sequencing primer. After gel electrophoresis the gel was transferred to wet blotting paper (Ford Golmedal), covered in Saran wrap, and dried in a BioRad Model 583 Gel Drier for 1 hour at 80°C .

10% Longranger™ Gels

10% Longranger™ gels were prepared and treated as above with the following exceptions: a flat toothed comb was used instead of a shark's-tooth comb and inserted toothed side first, 10ml Long Ranger™ were added to the gel mixture and the gel was run at a minimum temperature of 55°C to prevent hairpin formation.

SSCP Gel Electrophoresis

SSCP gels (16cm x 18cm x 1.5mm), were made and run in TBE gel buffer using a SE600 dual cooled vertical slab electrophoresis unit (Hoefer Scientific Instruments). The apparatus was assembled according to the manufacturer's instructions.

The gels were made by adding 25ml MDE Gel Solution to 12ml 5x TBE and the solution was made up to 100ml with distilled water. To the solution were added 40µl TEMED and 400µl freshly prepared AMPS. The solution was poured into the gel cell using a 25ml glass pipette. A 15-toothed flat edged comb was inserted, the top of the cell was covered with Saran wrap, and the gel left to polymerise for at least 1 hour. The gel was assembled in the apparatus and 30µl of each sample (15µl of Stop Solution & 15µl from restriction digest), were loaded onto the gel. The gel was run for 2-6 hours at a constant 200V. After electrophoresis the gel was stained using Silver Stain Plus Kit (BioRad).

Autoradiography

Autoradiography was carried out at room temperature in Cronex (DuPont) cassettes, using Cronex 4 (DuPont) X-ray film (30cm x 40cm). The exposure time depended on the strength of signal and varied from 1 day to 3 weeks. The films were developed in a X-OGRAPH Compact X2 automatic film processor.

Microbiological Methods

Bacterial Methods

Storage of Bacteria

For short term storage, bacteria were kept at 4°C on L-agar plates. Permanent stocks were prepared by adding 5 drops of sterile 100% glycerol to 1ml of a stationary phase culture in an Eppendorf. This was sealed with Parafilm and stored at -70°C.

Growth of Bacteria

Temporary stocks were generated by streaking permanent stocks to single colonies on L agar plates, which were incubated at 37°C overnight. Overnight cultures were grown by inoculating a single colony into L broth and shaking at 37°C.

Cosmid Cloning Methods

In Vitro Packaging of Cosmid DNA

Cosmids were *in Vitro* packaged by using the Gigapack[®] II XL Packaging Extract kit (Stratagene). The appropriate number of extract sets were removed from storage at -70°C and placed on dry ice. The sonic extract and freeze-thaw extract were thawed. 0.1-5 µg test DNA were added to the freeze-thaw extract at the point where the extract was just beginning to thaw. 15µl of the sonic extract were added, mixed by gentle pipetting, and the reaction was incubated at room temperature for 100 minutes. To terminate the reaction 500µl of SM buffer were added and the packaged DNA stored at 4°C.

Plasmid Methods

Maintenance of Plasmids

Plasmids were maintained by supplementing the media with ampicillin or when appropriate kanamycin.

Transformation

This method is derived from Mandel and Higa (1970). A fresh overnight culture was diluted 1 in 100 and grown at 37°C to an OD₆₅₀ of 0.4. The culture was chilled on ice and 25ml were centrifuged at 4.5krpm for 5 minutes at 4°C (MSE Centaur-2 bench centrifuge). The cell pellet was resuspended in 2ml ice-cold CaCl₂ and incubated on ice for 20 minutes. The cell suspension was then further centrifuged at 4.5krpm for 5 minutes at 4°C (MSE Centaur-2 bench centrifuge) and resuspended gently in 400µl ice-cold CaCl₂. 50-100ng of Plasmid DNA were added to 200µl of the competent cell suspension and incubated on ice for 30 minutes. The cells were heat-shocked for 2 minutes at 42°C followed by a 1 minute incubation on ice. 100µl of the transformed cells were spread on media (L-agar supplemented with ampicillin or ampicillin, IPTG and X-gal) and incubated overnight at 37°C.

Electroporation

This method is derived from Dower *et al.* (1988). A fresh overnight culture was diluted 1 in 100 (e.g. 10ml into 1 litre) and grown at 37°C to an OD₆₅₀ of 0.4-0.5. The culture was chilled on ice and centrifuged at 10krpm for 15 minutes at 4°C (Sorvall centrifuge, GSA rotor). The supernatant was removed and the pellet resuspended in an equal volume of ice-cold sterile 1mM HEPES (pH 7.0; made up with MilliQ water). The cell suspension was centrifuged at 10krpm for 15 minutes at 4°C (Sorvall centrifuge, GSA rotor). The supernatant was removed and the pellet resuspended in a half volume of ice-cold sterile 1mM HEPES (pH 7.0; MilliQ water).

The cell suspension was centrifuged at 10krpm for 15 minutes at 4°C (Sorvall centrifuge, GSA rotor). The supernatant was removed and the pellet resuspended in a 1/25 volume of ice-cold sterile 10% glycerol 1mM HEPES (pH 7.0). The cell suspension was centrifuged at 10krpm for 15 minutes at 4°C (Sorvall centrifuge, GSA rotor). The supernatant was removed and the pellet resuspended in a 1/50 volume of ice-cold sterile 10% glycerol 1mM HEPES (pH 7.0). The cell suspension was centrifuged at 10krpm for 15 minutes at 4°C (Sorvall centrifuge, SS34 rotor). The supernatant was removed and the pellet resuspended in a 0.08 volume of ice-cold sterile 10% glycerol 1mM HEPES (pH 7.0). 80µl aliquots of the electrocompetent were placed in Eppendorf tubes, frozen on dry ice and stored at -70°C.

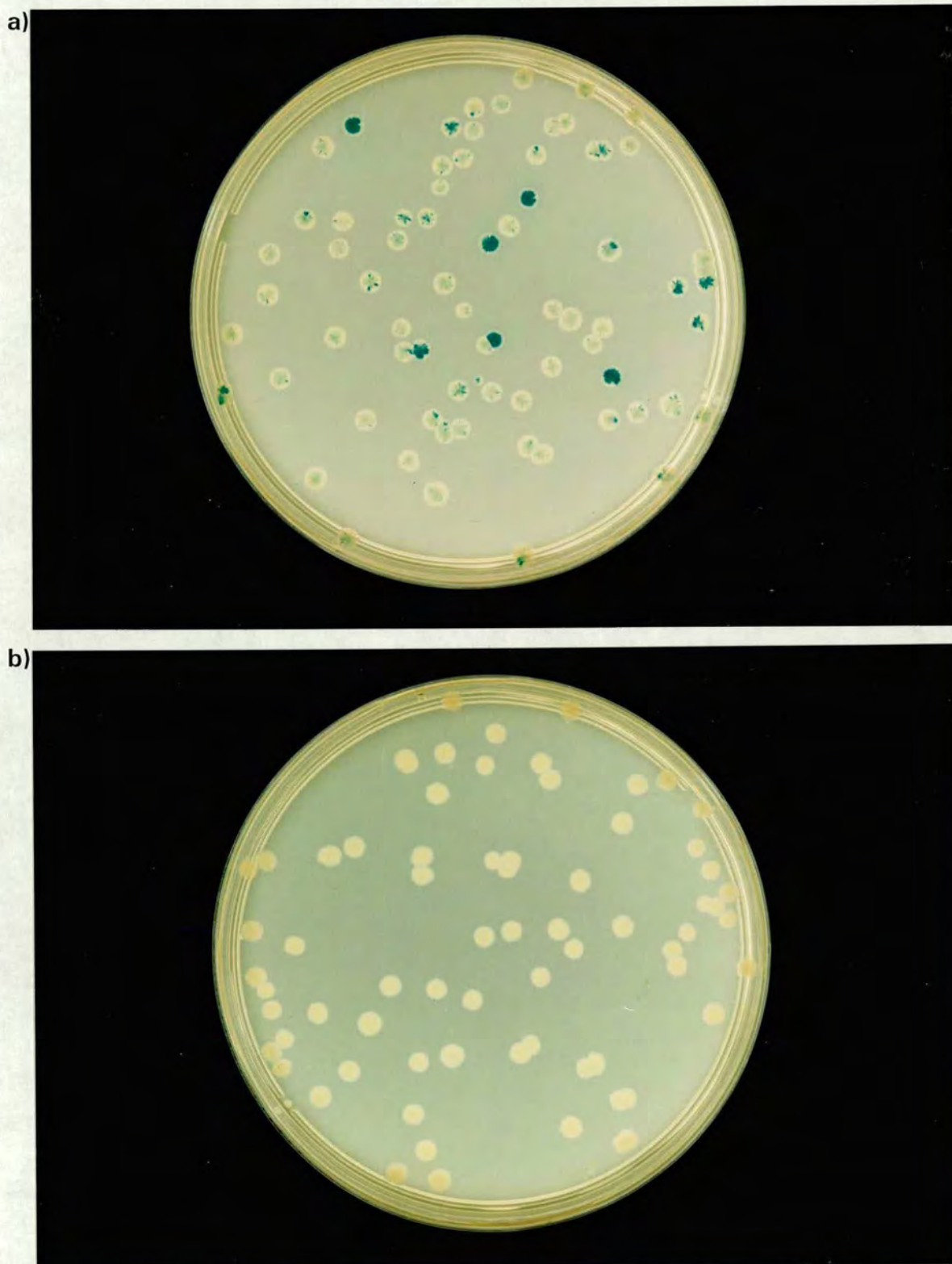
To transform, 10-50ng of plasmid DNA were added to 40µl electrocompetent cells in an electroporation cuvette (1mm electrode gap) on ice. The cuvette was placed in a Gene Pulser II apparatus (BioRad) and pulsed at 2.5kV, 25µF at a resistance of 200Ω. 1 ml SOC broth was added immediately and the cell suspension transferred to a 5ml plastic tube. After incubation at 37°C for 1 hour, 200µl were spread on media (L-agar supplemented with ampicillin or ampicillin, IPTG and X-gal), and incubated overnight at 37°C.

CHAPTER 3:

The Search for the 'Mutation' in pMS7

Figure 3.1

Colony morphology of DL733 containing a) pAC2 and b) pMS7 when grown on AXI media.



Introduction

DNA palindromes longer than 150-200bp can not be cloned in wild type *E. coli*, either the replicon containing the palindrome is so poorly replicated that it is inviable (Warren & Green, 1985; Yoshimura et al., 1986), or the palindrome is so unstable that it suffers partial or complete deletion (Collins, 1981; Hagen & Warren, 1983). These phenomena have been termed inviability and instability respectively. The propagation of palindromic sequences is affected by the genotype of the host. *E. coli sbcCD* mutants are known to allow the replication of DNA containing long palindromes (Collins *et al.*, 1982; Chalker *et al.*, 1988; Gibson *et al.*, 1992).

Before commencement of this work three plasmids, pAC2, pMS5 and pMS7 were constructed, or isolated by other workers (Chalker & Shaw, unpublished). pAC2 was constructed by ligating a 571bp imperfect palindrome into the *EcoRI* site of pUC18. DL733 ($\Delta sbcCD::Kan^r$) was transformed with pAC2 DNA and the clone chosen had a characteristic phenotype on L-agar supplemented with Ampicillin, X-gal and IPTG (AXI media), white colonies with dark blue sectors (see Fig 3.1a). pMS5 and pMS7 were both isolated as derivatives of pAC2 which have different phenotypes on AXI. (see Fig.1.3b for the phenotype of pMS7). pMS5 has a similar phenotype on AXI to that of pMS7, only with more sectoring.

pUC plasmids were first developed by Vieira and Messing (1982). They contain a gene which codes for the α subunit of β -galactosidase (*lacZ'*). Cloning in pUC18 involves insertional inactivation of the *lacZ'*. The enzyme β -galactosidase, whose synthesis can be induced by IPTG, can hydrolyse the chromogenic substrate (X-gal), to give a dark blue colour (Horwitz *et al.*, 1964). *E. coli* hosts, encoding a defective form of β -galactosidase may be complemented by transformation with pUC18 and will give rise to blue colonies. Insertion into this gene leads to the production of an inactive α subunit. Consequently, *E. coli* harbouring recombinant plasmids will give rise to white colonies. This provides a simple histochemical test, to distinguish parental clones from recombinants.

With consideration to the above, the most ~~reasonable~~ conclusion concerning the phenotype of pAC2, is that the sectorised phenotype (as seen in Fig. 3.1a), is due to instability of the 571bp palindrome. This conclusion was assumed at the beginning of this work. The difference seen between the phenotype of pAC2, pMS5 and pMS7 was explained by the presence of a mutation in pMS5 and pMS7 which stabilised the propagation of palindromes (Leach & Blake, unpublished). At the onset of this work the aim was to construct a deletion resistant cosmid using pM* (pMS7 with the 571bp palindrome removed, see Fig 3.2), as part of the backbone. The concept was, if a 'mutation' could stabilise a palindrome which appeared unstable, then the 'mutation' could perhaps stabilise other unstable sequences (see Chapter 1 of this thesis).

Sequencing of the 'mutant' plasmids revealed no gross change in their sequence compared to that of pUC18 (Blake, 1996). In this chapter, are described the results of two other methods, Single-Strand Conformation Polymorphism analysis (SSCP; Orita *et al.*, 1989) and Fragment Swap analysis, which were used to try to determine the position of the mutation in pM* (for a summary of these methods see Fig. 3.3 and Fig. 3.4).

Results

SSCP analysis

A series of different single-stranded fragments were run on SSCP polyacrylamide gels. These included fragments generated by digestion of pUC18 and pM* with: *Aat*II, *Apa*LI, *Bgl*II, *Bst*II, *Dpn*I, *Mbo*I, *Pvu*II and *Taq*I. The fragments were visualised by silver staining using Silver Stain Plus Kit (BioRad).

No differences in strand mobility, between fragments generated from pUC18 and pM* were observed (data not shown).

Fragment Swap Experiments

A set of four fragment swaps were performed to try to determine the approximate position of the 'mutation'. Each experiment involved: a) restriction of the plasmids with two endonucleases each of which cuts the backbone once, b) gel purification of the fragments, c) ligation of the fragments and d) transformation of DL733 ($\Delta sbcCD::Kan^r$) with the ligation mixes and the controls. Two -ligase controls were included to estimate the relative amount of uncut background, and three +ligase controls of each of the individual fragments were included to measure the background from each fragment. The results from the four experiments are in Tables 3.1, 3.2, 3.3 and 3.4. The controls are shown as percentage totals in each set of data. The actual numbers of each colony type in the control experiments were very low.

Note, *BsrFI* is in the ampicillin resistance gene of pUC18. Other fragment swaps not involving the *BsrFI* site were attempted but, these gave a large number of white clones. This was probably due to junk DNA, ligated in the fragment swap. Fragment swaps involving the *BsrFI* site alleviated the problem of background 'whites' (see Fig. 3.2 for position of endonuclease sites).

Figure 3.2

An overview of the chronological construction of pAC2, pMS7 and pM*. An approximate restriction map of the endonucleases used in the fragment swap experiments is also shown.

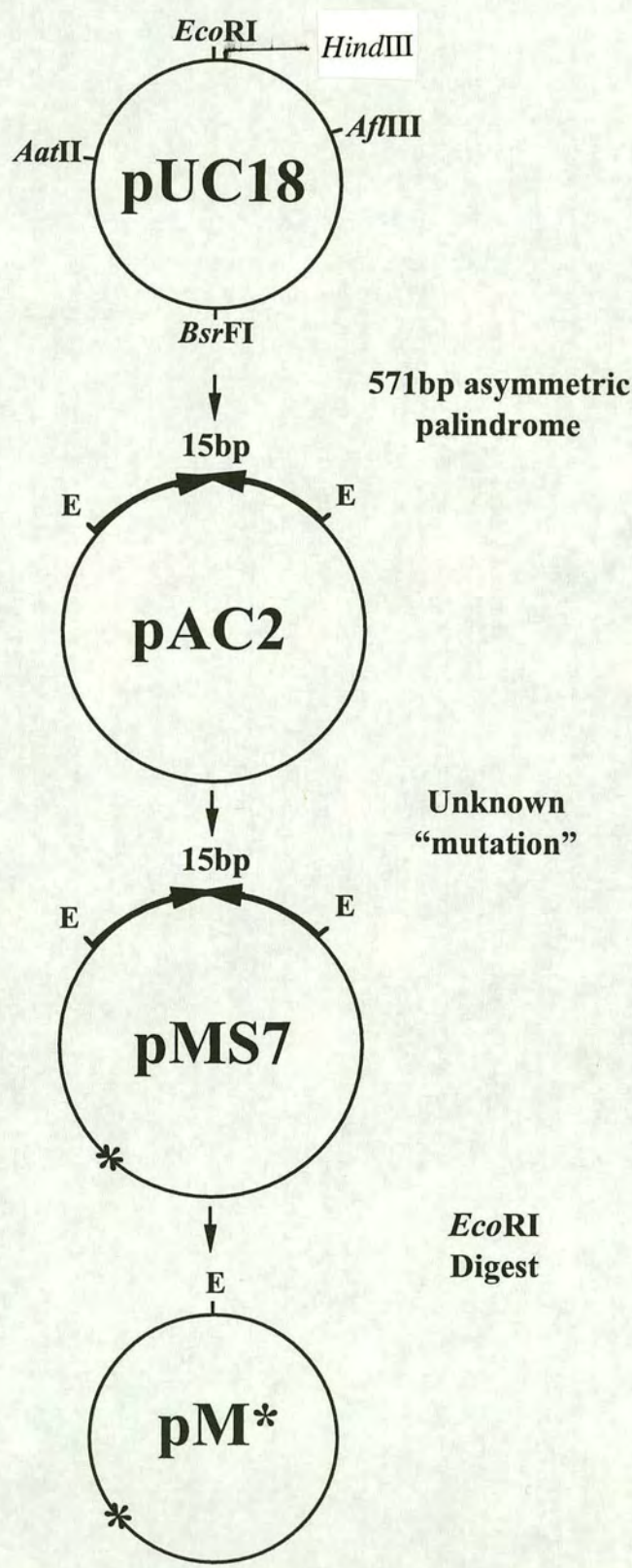


Figure 3.3

The principle of SSCP analysis: a) plasmids cut with restriction endonucleases, to generate double stranded fragments, b) fragments chemically denatured, single strands fold in a different way, and c) native gel electrophoresis reveals any differences.

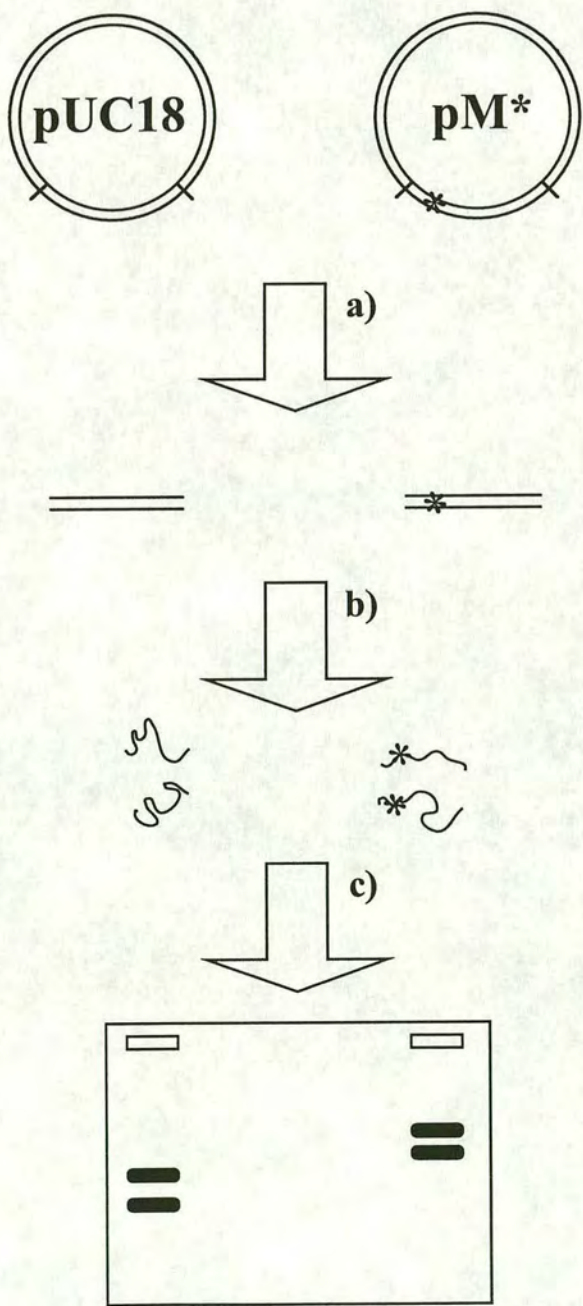
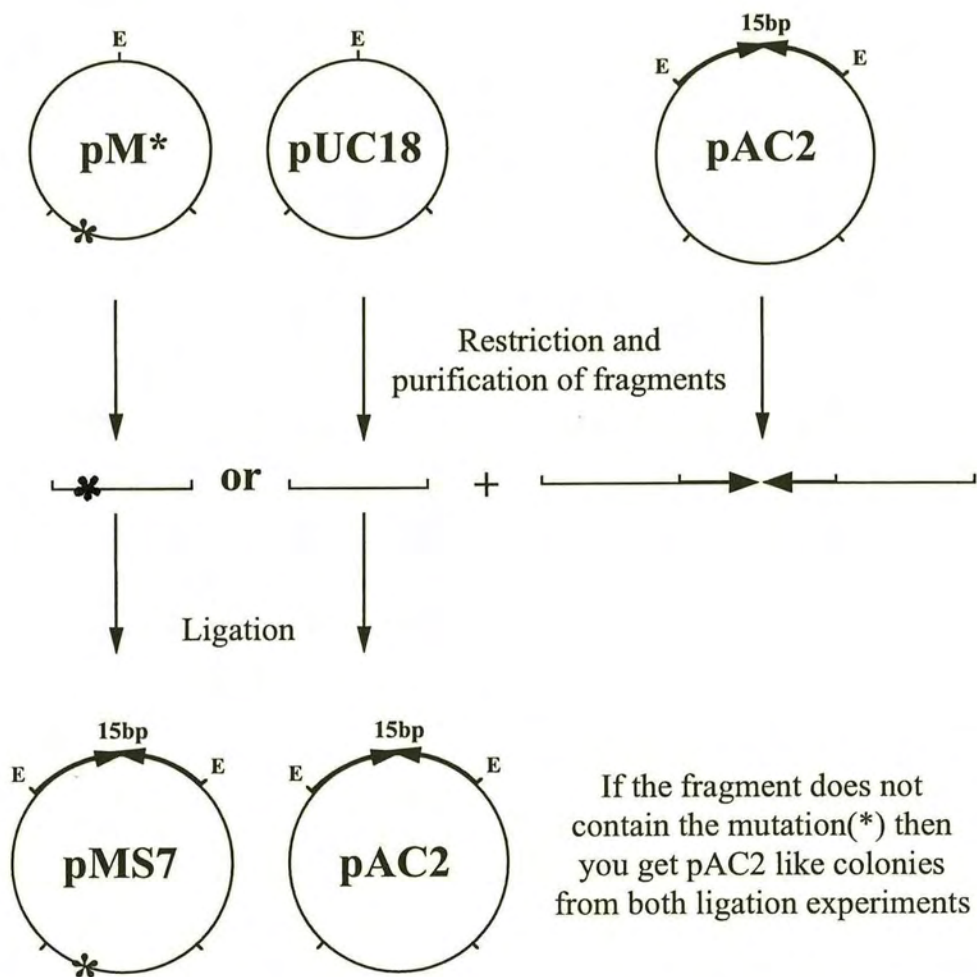
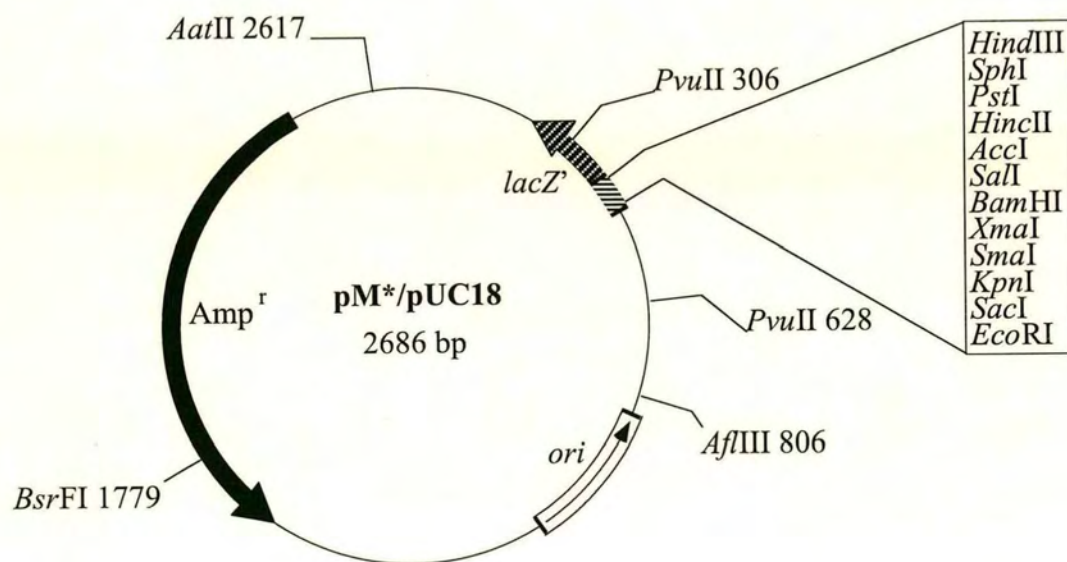


Figure 3.4

Summary of the principles of the Fragment Swap experiments.





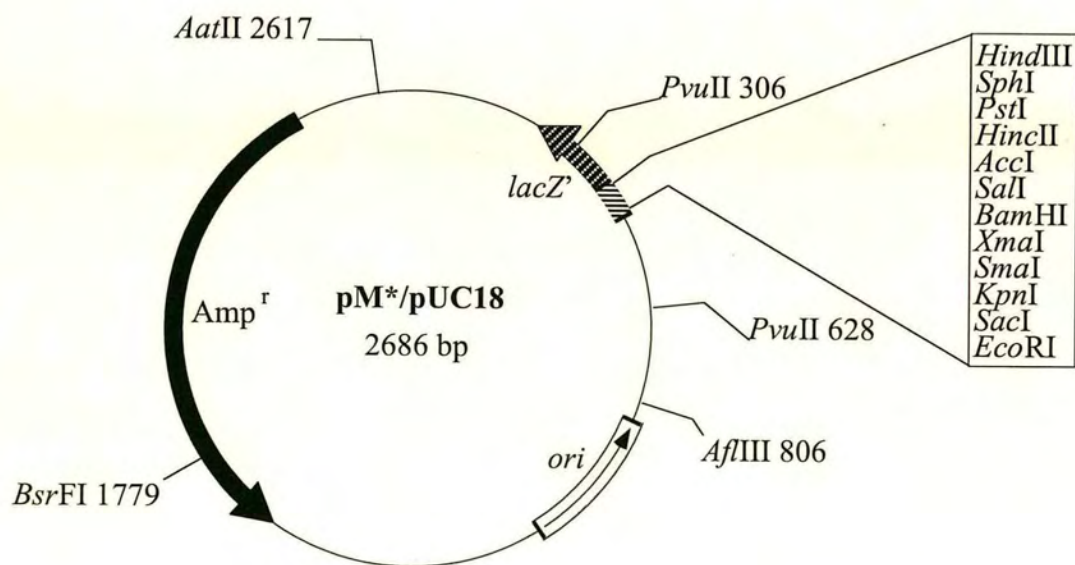
Restriction diagram of pUC18 (or pM*), showing the relevant sites for the table on the facing page. pAC2 and pMS7 have a 571bp palindrome ligated into the *Eco*RI site.

Table 3.1

Percentage colony type after fragment swap 1: (*BsrFI* and *AflIII*).

Fragment Swap Ligation Mix	% Total			TOTAL NUMBER OF COLONIES
	Blue colonies	Sectoring colonies "pAC2 like"	White colonies	
pUC18/pAC2 + ligase	34	66	-	902
pM*/pAC2 + ligase	18	82	-	197
pUC18/pAC2 - ligase	-	-	-	0
pM*/pAC2 - ligase	-	-	-	0
pUC18 fragment + ligase	-	-	-	0
pAC2 fragment + ligase	100	-	-	40
pM* fragment + ligase	-	-	-	0

pM*, pUC18 and pAC2 were digested to completion with *BsrFI* and *AflIII*. The 1kb fragments from pM* and pUC18, were gel purified and ligated to the 2.3kb fragment from pAC2, which also was gel purified. DNA was used to transform DL733.



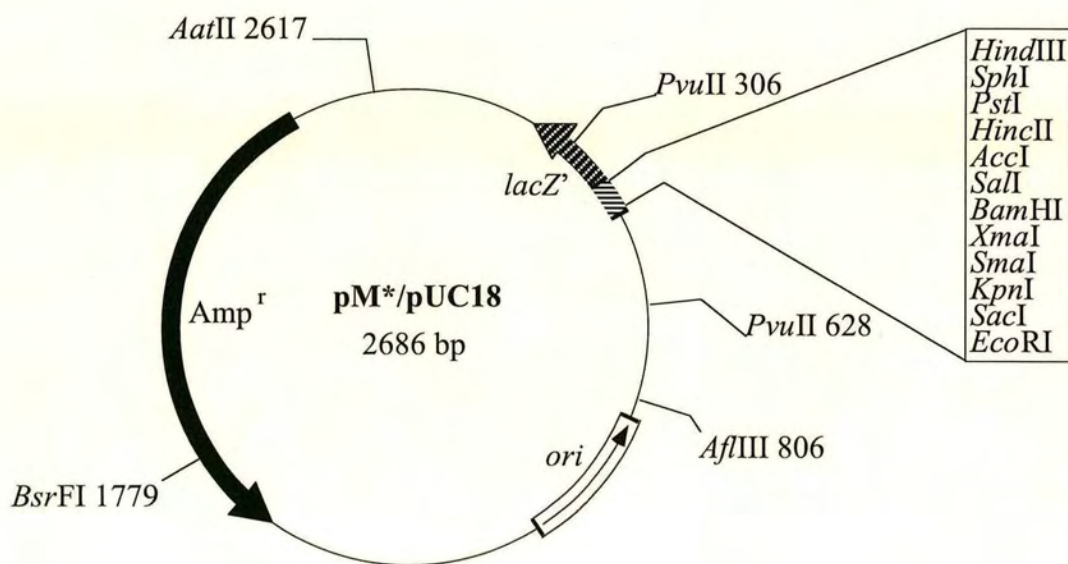
Restriction diagram of pUC18 (or pM*), showing the relevant sites for the table on the facing page. pAC2 and pMS7 have a 571bp palindrome ligated into the *EcoRI* site.

Table 3.2

Percentage colony type after fragment swap 2: (*Bsr*FI and *Afl*III).

Fragment Swap Ligation Mix	% Total			TOTAL NUMBER OF COLONIES
	Blue colonies	Sectoring colonies “pAC2 like”	White colonies	
pAC2/pUC18 + ligase	13	67	20	402
pMS7/pUC18 + ligase	41	-	59	70
pAC2/pUC18 - ligase	93	5	2	41
pMS7/pUC18 - ligase	100	-	-	32
pUC18 fragment + ligase	100	-	-	6
pAC2 fragment + ligase	36	44	20	85
pM* fragment + ligase	100	-	-	9

pUC18, pAC2 and pMS7 were digested to completion with *Bsr*FI and *Afl*III. The 2.3kb fragments from pAC2 and pMS7, were gel purified and ligated to the 1kb fragment from pUC18, which also was gel purified. DNA was used to transform DL733.



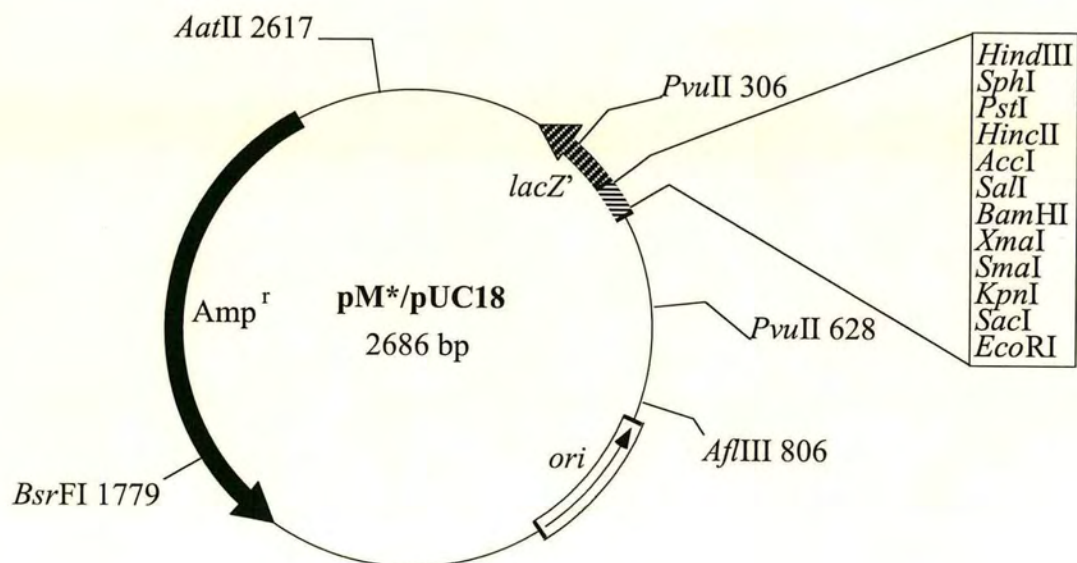
Restriction diagram of pUC18 (or pM*), showing the relevant sites for the table on the facing page. pAC2 and pMS7 have a 571bp palindrome ligated into the *EcoRI* site.

Table 3.3

Percentage colony type after fragment swap 3:(*Bsr*FI and *Aat*II).

Fragment Swap Ligation Mix	% Total			TOTAL NUMBER OF COLONIES
	Blue colonies	Sectoring colonies “pAC2 like”	White colonies	
pUC18/pAC2 + ligase	8	78	15	89
pM*/pAC2 + ligase	6	84	10	153
pUC18/pAC2 - ligase	-	-	-	0
pM*/pAC2 - ligase	-	100	-	1
pUC18 fragment + ligase	-	-	-	0
pAC2 fragment + ligase	-	100	-	3
pM* fragment + ligase	100	-	-	3

pM*, pUC18 and pAC2 were digested to completion with *Bsr*FI and *Aat*II. The 0.8kb fragments from pM* and pUC18 were gel purified and ligated to the 2.4kb pAC2 fragment, which also was gel purified. DNA was used to transform DL733.



Restriction diagram of pUC18 (or pM*), showing the relevant sites for the table on the facing page. pAC2 and pMS7 have a 571bp palindrome ligated into the *EcoRI* site.

Table 3.4Percentage colony type after fragment swap 4: (*Bsr*FI and *Hind*III).

Fragment Swap Ligation Mix	% Total			TOTAL NUMBER OF COLONIES
	Blue colonies	Sectoring colonies "pAC2 like"	White colonies	
pUC18/pAC2 + ligase	25	39	36	2790
pM*/pAC2 + ligase	6	45	49	3690
pUC18/pAC2 - ligase	-	-	-	0
pM*/pAC2 - ligase	-	-	-	0
pUC18 fragment + ligase	100	-	-	1084
pAC2 fragment + ligase	5	51	44	79
pM* fragment + ligase	79	8	13	96

pM*, pUC18 and pAC2 were digested to completion with *Bsr*FI and *Hind*III. The 1.3kb fragments from pM* and pUC18 were gel purified and ligated to the 2.0kb pAC2 fragment, which also was gel purified. DNA was used to transform DL733.

Conclusion

SSCP analysis

The SSCP experiments did not reveal any strand mobility differences between fragments generated from pM* and pUC18. This presents two possibilities: a) the assay was not sensitive enough to detect differences between pM* and pUC18 fragments, and b) there are no differences between pM* and pUC18, and the ‘mutation’ is not present in pM*.

Fragment Swap Experiments

The results presented above seem to suggest a possible difference between pM* and pUC18. In the first fragment swap the *BsrFI/AflIII* 1kb region of pM* and pUC18 were ligated to the 2.3kb fragment of pAC2. If the ‘mutation’ were present in the 1kb region then the swap would have generated sectoring colonies with the pUC18 swap and white colonies with the pM* swap. The data in Table 3.1 shows that both the pUC18/pAC2 and pM*/pAC2 swaps generated sectoring colonies. This suggests that the ‘mutation’ is not present in the 1kb region of pM*.

In the second fragment swap *BsrFI/AflIII* 2.3kb fragments from pAC2 and pMS7 were ligated to the 1kb fragment from pUC18 (the converse of fragment swap 1). This swap is not a direct comparison of pM* and pUC18, but it was necessary to use pAC2 and pMS7, as pM* does not contain the palindrome. The data shown in Table 3.2 shows that the pMS7/pUC18 swap did not generate sectoring colonies. The absence of sectoring colonies from this swap strongly suggests that the 2.3kb section of pMS7 contains a ‘mutation’ which appears to stabilise palindromes. The generation of white colonies in the pAC2/pUC18 swap could not be explained, but it is possible that the experiment somehow enriched a sub-population present in pAC2.

In the third fragment swap *BsrFI/AatII* 0.8kb fragments from pM* and pUC18 were ligated to the 2.4kb fragment of pAC2. The data shown in Table 3.3

shows that the pUC18/pAC2 and pM*/pAC2 swaps generated both sectoring colonies and white colonies. The generation of sectoring colonies in the pM*/pAC2 swap suggests the 'mutation' is not present in the 0.8kb region of pM*. Taking the data from fragment swap 1 and fragment swap 3 the 1.8kb region of pM* between *Afl*III and *Aat*II does not appear to contain the 'mutation'.

The fourth fragment swap was designed to increase the resolution of the experiment. *Bsr*FI/*Hind*III 1.3kb fragments from pM* and pUC18 were ligated to the 2.0kb fragment of pAC2. The data shown in Table 3.4 shows the generation of sectoring colonies in both the pUC18/pAC2 and pM*/pAC2 swaps. The presence of sectoring colonies in this swap suggests the 'mutation' is not in the region between the *Bsr*FI and *Hind*III sites. This result taken together with all the previous data suggests, by elimination, that the 'mutation' is present in the 398bp fragment between *Hind*III and *Afl*III of pMS7.

It is not possible to draw any direct conclusions from the results presented above. When observed on their own, they seem to suggest a fragment which may contain a 'mutation'. The results obtained were confused, by the presence of large numbers of white colonies in the pAC2/pUC18 fragment swap, in experiment two and the pUC18/pAC2 fragment swap, in experiments three and four. If the generation of white colonies were due to the presence of a 'mutation', then white colonies should not be generated by these swaps. A possible alternative for their appearance is enrichment of a sub-population due to the processes involved in the experiment.

NOTE: I have discussed the data above as it stands on its own. Further data presented in Chapter 5, have permitted an alternative conclusion regarding the fragment swap experiments.

CHAPTER 4:

The cloning of DNA palindromes

Introduction

As described in the introduction of this thesis, the DNA sequence of a palindrome plays an important role in its genetic stability. The kinetics and energetics of cruciform extrusion (reviewed by Murchie & Lilley, 1992) are primarily determined by the sequence of the palindrome and the level of DNA supercoiling. DNA palindromes are believed to extrude to form cruciform structures by two pathways, S-type extrusion and C-type extrusion. S-type extrusion is characterised by a requirement for cations, relatively moderate activation entropies and enthalpies, and a profound effect of central sequence changes on the reaction kinetics. It involves the melting of a small (~10bp) region at the centre of a DNA palindrome in negatively supercoiled DNA. The single stranded region generated, intrastrand base pairs to form a protocruciform structure. Branch migration, driven by DNA-supercoiling, leads to the extrusion of the full length cruciform. Therefore, the sequence at the centre of a palindrome, which extrudes by an S-type mechanism is important in determining its stability when cloned. It is known that AT-rich centres have the lowest kinetic barrier to cruciform extrusion *in vivo* (reviewed by Davison, 1994). Asymmetry at the palindrome's centre can also alleviate instability and inviability (Collins *et al.*, 1982; Warren & Green, 1985).

pMS7 appears to stabilise the propagation of a 571bp imperfect palindrome (Leach & Blake, unpublished). A priority was to demonstrate that the enhanced stability was not specific to the 571bp palindrome. DNA from four λ phage, each containing different palindromes was provided by Angus Davison. For details see Davison (1994). λ DRL180 contains an imperfect palindrome with ten base pairs of asymmetry. λ DRL167, λ DRL170 and λ DRL171 contain perfect palindromes. In this chapter, are described the results from cloning these four DNA palindromes with different central sequences in pM*.

Results

Cloning Other DNA Palindromes

In order to test if it was possible to stabilise other DNA palindromes in pM*, four palindromes each with different central sequences, were gel purified from the bacteriophage DNA after *EcoRI* digestion and were ligated separately into the *EcoRI* site of pM* (see Fig 3.2 for how pM* was isolated). Each of these palindromes was about 460bp long. A summary of the palindromes used in this experiment may be seen in Table 4.1. As a control the four palindromes were ligated into the *EcoRI* site of pUC18 also.

Table 4.1

Summary of the palindromes used.

λ Phage [†]	Restriction site	Sequence at centre	Plasmid name
λ DRL180	<i>XhoI</i>	GTGGATCTCA → ← 10bp	pM*180
λ DRL167	<i>SacI</i>	GAGCTC → ←	pM*167
λ DRL170	<i>SphI</i>	GCATGC → ←	pM*170
λ DRL171	<i>AatII</i>	GACGTC → ←	pM*171

[†]Lambda phage DNA was provided by Angus Davison (see Davison, 1994).

It was possible, to clone all four palindromic sequences in pM*, but not in pUC18. The presence of the palindromes in pM* was confirmed by agarose gel electrophoresis of an *EcoRI* and *HindIII* digest of each of the pM* vectors. An *EcoRI* or *HindIII* digest of pM*180 and pM*167 may be seen in Fig. 4.1. Fig. 4.2 shows

pM*180 has less sectoring than pM*167. pM*180 therefore appears to be more stable than pM*167.

Lowered DNA Yield/Copy Number

DNA was prepared using a QIAGEN Plasmid Midi Kit. When preparing DNA, a lower DNA yield than was usual was observed. The yield seemed to be about 10% of normal (normal=2-5µg/ 10 ml culture). This led to an investigation of copy number of pM*180 and pM*167 compared to the copy number of pUC18. *E. coli* DL733 harbouring pM*167, pM*180 and pUC18 were plated on AXI media containing increasing concentrations of ampicillin. The number of colonies on control plates containing no ampicillin was taken as 100% survival. As Fig. 4.3 illustrates pM*180 and pM*167 had a much lowered copy number compared to that of pUC18. At concentrations of about 400µg/ml ampicillin the colonies surviving were blue and therefore had deleted the palindrome. Using Fig 4.3 it possible to estimate the concentration at which there is 50% survival. pM*167 has 50% survival at an ampicillin concentration of approximately 50µg/ml, while pM*180 has 50% survival at an ampicillin concentration of approximately 200µg/ml. Therefore, pM*167 appears to have a lower copy number than pM*180.

NOTE: An attempt was made to increase the ratio of the palindrome-containing DNA in the plasmid DNA isolated. pM*167 DNA was used to transform *E. coli* DL901 (*sbC*, $\Delta pcnB::Kan^r$). This experiment did not give the desired result and exaggerated the low plasmid DNA yield.

Figure 4.1

Negative picture of ethidium bromide stained, agarose gel: a) *Eco*RI digest of pUC18, pM*, pMS7, pM*180 & pM*167, b) *Hind*III digest of pMS7, pM*180 & pM*167.

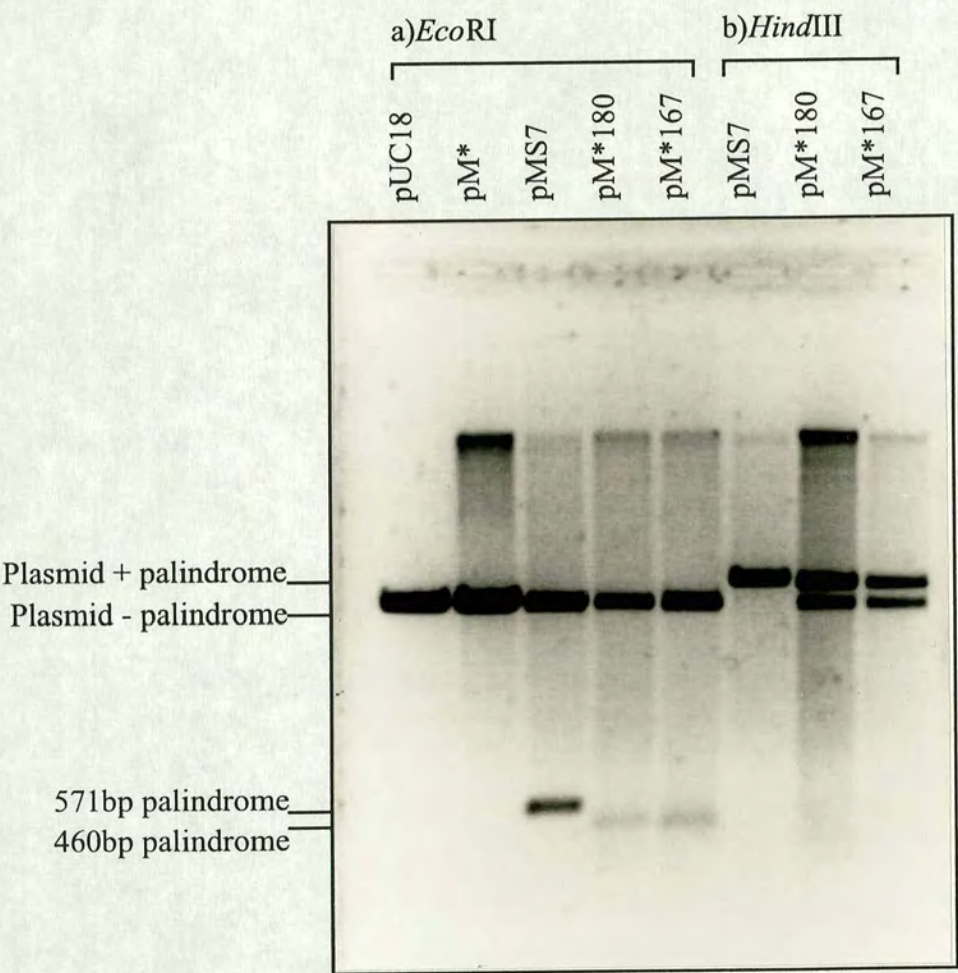


Figure 4.2

Colony morphology of DL733 containing a)pM*180 and b) pM*167, when grown on AXI media.

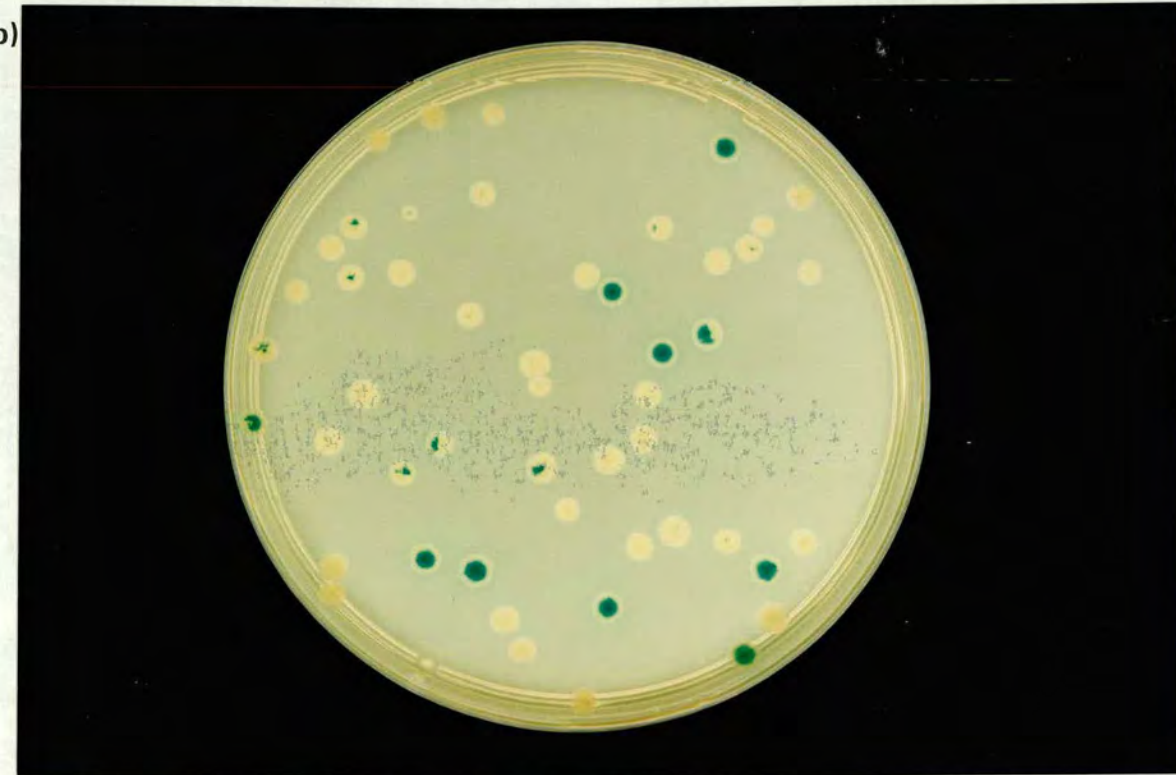
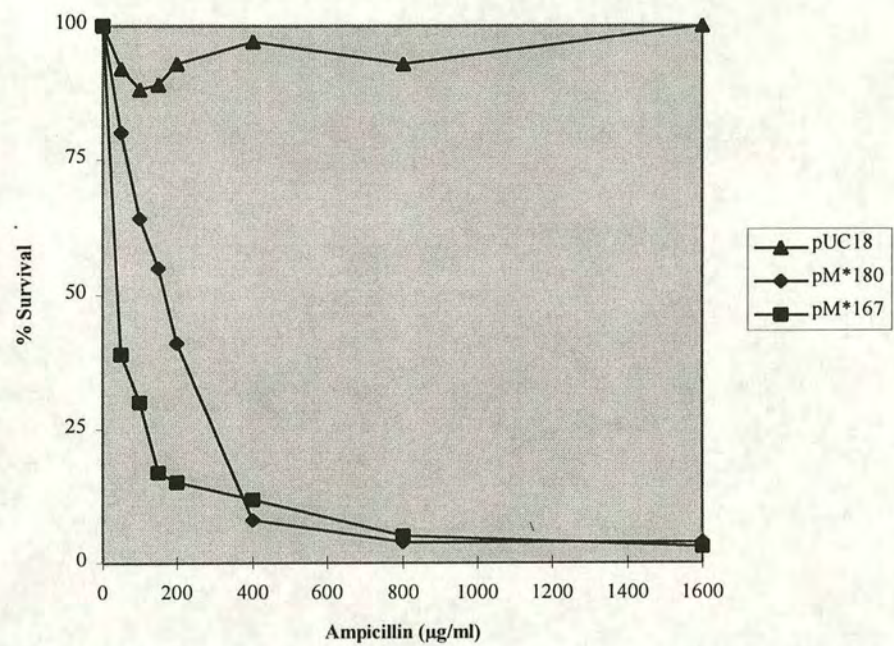


Figure 4.3

Average survival on Ampicillin: (N=20)



E. coli DL733 harbouring pM*167, pM*180 and pUC18 were plated on AXI media containing increasing concentrations of ampicillin. The number of colonies on control plates containing no ampicillin was taken as 100% survival.

Conclusion

Cloning Other DNA Palindromes

The purpose of the cloning experiments described above was to demonstrate a difference between pM* and pUC18. As discussed in the introduction to Chapter 3, pM* was derived from pMS7, and thought to contain a 'mutation' which permitted the propagation of the 571bp palindrome. The results suggested that it was possible to clone the four new DNA palindromes in pM*, but not in pUC18.

NOTE: After the conclusion of these cloning experiments, I cloned the original 571bp palindrome into pUC18. The colony morphology of the majority of the clones obtained were similar to that of pMS7 (white colonies with no sectors). This contradicted the conclusion that pMS7 had a 'mutation' which helped to stabilise the propagation of palindromes.

Therefore, it was not possible to conclude if the 'mutation' in pMS7 and pM* existed or not. At this stage of investigation, some evidence suggested that pM* was identical to pUC18, and that the mutation did not exist (sequencing of pM*, Blake, 1996; and SSCP analysis as reported in Chapter 3 of this work). However, the fragment swap analysis in Chapter 3 and the cloning experiments in this Chapter suggested that there was a difference between pM* and pUC18.

Lowered DNA Yield/Copy Number

When preparing pM*180, pM*167, pM*170 and pM*171 DNA, a low DNA yield was observed. Qualitative analysis of the DNA prepared, revealed pM*180 DNA consisted of about 50% palindrome containing pM*180, and 50% pM* backbone (as illustrated in Fig. 4.1). pM*167, pM*170 and pM*171 DNA were all found to contain low ratios of vector plus palindrome (Fig. 4.1 and other data not shown). Inviability and/or instability problems associated with the cloning of long DNA palindromes (reviewed by Leach, 1994), provides an explanation of the low DNA yields obtained. The palindromes cloned may be so unstable that they delete

out from the plasmid backbone at a high frequency. Or, if they form a stable cruciform secondary structure they would have the potential to terminally block DNA replication resulting in an inviable replicon. Inviability and instability are likely to both play a part in the low yield observed in these cloning experiments.

Imperfections introduced into perfect palindromes are known to alleviate problems of instability and/or inviability (Collins *et al.*, 1982; Warren & Green, 1985). This provides an explanation for the relative stability of pM*180, as the palindrome in pM*180 contains 10bp of asymmetry.

The difference in the stability of pM*180 compared to the other plasmids was supported with the investigation of copy number (see Fig. 4.3). pM*180 and pM*167 had a much lower copy number than that of pUC18. pM*167 had the lowest copy number with a 50% survival at an ampicillin concentration of 50µg/ml. The results from the copy number experiments, confirm the conclusion that palindromes interfere with DNA replication by forming secondary structure, and in this case lower the number of viable replicons within the host cell.

CHAPTER 5:

The isolation of a novel mutation: DIR

Introduction

Sequences which have the potential to form unusual secondary structures are known to promote pausing of DNA replication and can stimulate replication slippage (Sinden & Wells, 1992). Such sequences include: a). oligopurine-oligopyrimidine sequences, able to adopt a triple helix structure in supercoiled DNA (H-DNA), both *in vivo* (Rao *et al.*, 1988) and *in vitro* (Dayn *et al.*, 1992; Samadashwily *et al.*, 1993), b). tri- and tetra-nucleotide repeats, thought to form pseudo-hairpins when single stranded (Wells, 1996; Richards & Sutherland, 1992), and recently shown to have the potential to form a K^+ -dependent quadruplex secondary structure (Usdin & Woodford, 1995), c). alternating GC or GT dinucleotide repeats (microsatellite repeats), capable of forming Z-DNA (Hite *et al.*, 1996; Freund *et al.*, 1989; Klysik *et al.*, 1982), and d). DNA palindromes (perfect inverted repeats containing no spacer region at the central sequence), able to form hairpins, when single stranded, or cruciforms, when double stranded (for a review see Leach, 1994).

pAC2 has a very different phenotype to that of pMS7 (as seen in Chapter 3, Fig 3.1). The most logical conclusion concerning the phenotype of pAC2, is that the sectorised phenotype is due to instability of the 571bp palindrome. This conclusion was assumed at the beginning of this work. The difference seen between the phenotype of pAC2, pMS5 and pMS7 was explained by the presence of a mutation in pMS5 and pMS7 which stabilised the propagation of palindromes (Leach & Blake, unpublished). Blake (1996), demonstrated that the simplest multimeric form of pAC2, was a heterodimer in which only one copy of the palindrome was present and that the simplest form of pMS7 was a homodimer with the palindrome present in both copies of the plasmid. This presents an alternative explanation for the difference of phenotype observed between pAC2 and pMS7, that colonies harbouring pAC2 DNA sectorised because pAC2 was a heterodimer.

In this chapter, are described the results of an experiment which is a continuation of the fragment swap experiments described in Chapter 3, and the subsequent discovery, isolation, and characterisation of a novel type of mutation

which I have called DIR (Direct and Inverted Repeat). The mutation appears to be caused primarily by replication slippage and to be stimulated by the presence of the 571bp asymmetric DNA palindrome.

Results

Cut and Ligate experiment

pAC2 and pMS7 were digested with *Bsr*FI, re-ligated and the ligation mix DNA was used to transform DL733. The results are summarised in Table 5.1. This experiment was designed to investigate the effects of the processes in the fragment swap experiments described in Chapter 3. pAC2 and pMS7 were known to exist as dimers (Blake, 1996), so this experiment may also be viewed as the monomerization of pAC2 and pMS7.

The results in Table 5.1 indicate an increase in the yield of highly sectoring colonies derived from pMS7; from 1% to 8.5%. They also indicate changes in the ratios of colony type after pAC2 has been cut and religated. Plasmid DNA was isolated from a large number of sectoring “pAC2 like” clones (derived from both pAC2 and pMS7). Restriction analysis revealed that although they had the sectoring phenotype of pAC2 they did not contain the 571bp DNA palindrome. One such plasmid (denoted pDIR1) derived from the monomerisation of pAC2 was chosen for further analysis. Restriction analysis with *Pvu*II suggested that the plasmid was predominantly a heterodimer of pUC18, and pUC18 with an insert of approximately forty nucleotides (see Fig 5.1). Furthermore, this analysis suggested that pAC2 heterodimers contained the 571bp palindrome on one side and this same forty nucleotides insert on the other. This also revealed that pMS7 was primarily present as a homo-multimer.

Fine-Structure Analysis of pDIR1

pAC2, pDIR1, pMS7 and pUC18 DNA were digested with *Eco*RI and radiolabelled using [α - 35 S] dATP and Klenow enzyme. A sample of the purified DNA was digested with *Bam*HI, run on a 10% Longranger™ polyacrylamide gel and bands were visualised by autoradiography (see Fig 5.2). The restriction pattern obtained confirmed the presence of a second *Eco*RI site in the polylinker region, giving a 46bp fragment (42bp plus 4bp *Eco*RI overhang) in both pAC2 and pDIR1. pMS7 and pUC18 did not have this second *Eco*RI site. In contrast all four plasmids have the correct 25bp (21bp plus 4bp) *Eco*RI to *Bam*HI fragment.

Sequencing of the 42bp Insert Present in pDIR1

DNA sequencing of the polylinker region of pDIR1, pUC18 and pAC2 was carried out on both strands using two primers: 576L and -40 (see Fig 5.3). (-40 is a standard M13 sequencing primer, available commercially).

Figure 5.3

Primers for the sequencing of the polylinker region of pUC18.

(576L) 5`-**GACTGGAAAGCGGGCA**-3`

(-40) 5`-**GTTTTCCCAGTCACGAC**-3`

Since pDIR1 and pAC2 were found to be predominantly hetero-multimers, sequencing of the 42bp insert required its separation from the other component of the multimer. This was achieved in two ways. Firstly, the *Pvu*II fragment containing the insert was purified away from the other parts of the multimer by agarose gel electrophoresis and sequencing was performed on the DNA fragment isolated from the gel. Secondly, pDIR1 DNA was used to transform JM83 and a homo-multimer containing the 42bp insert was isolated that could be sequenced directly. Normal

DNA sequencing (Sanger, 1977), revealed a region in the multicloning site where the DNA polymerase seemed to stall. This suggested the presence of secondary structure. Sequencing in the presence of 10% DMSO destabilised the secondary structure, allowing the polymerase to sequence through most of this region and revealed the presence of a novel mutation (see Fig 5.4). The sequence shows the presence of a duplicated *EcoRI* to *BamHI* multi-cloning site fragment with the same region inverted between the duplication.

Analysis of pDIR1 derivatives

When *E. coli* cells containing pDIR1 are plated on AXI medium, two types of colony phenotype are observed. One type is sectorial and the others are completely blue, similar to cells containing pUC18. pDIR1 DNA was used to transform JM83, DL733 and DL795 and Plasmid DNA was isolated from the blue and sectorial colonies formed. Agarose gel electrophoresis of uncut DNA (Fig. 5.5) revealed that plasmid DNA isolated from sectorial colonies with JM83 (wild type) and DL733 ($\Delta sbcCD::Kan^r$) genetic backgrounds were high multimeric forms. The plasmid DNA isolated from blue colonies with JM83 and DL733 genetic backgrounds were mostly monomeric. DL795 (*recA*⁻, *sbcC*⁻), does not contain the mutation for blue/white selection in pUC plasmids. Two plasmid populations, one mostly present as monomers and the other mostly present as multimeric forms were isolated from the DL795 genetic background (see Fig. 5.5, lanes 11 and 13).

Sequencing and restriction with *PvuII* of plasmid DNA isolated from blue colonies revealed that they did not contain the DIR structure and were really pUC18 (data not shown).

NOTE: The homo-multimer plasmid containing the 42bp insert that could be sequenced directly, is in lane 3 in Fig. 5.5.

Table 5.1Percentage colony type, after digestion with *Bsr*FI, ligation and transformation.

Plasmid	% Total			TOTAL NUMBER OF COLONIES
	Blue colonies	Sectoring colonies “pAC2 like”	White colonies	
pAC2 cut with <i>Bsr</i> FI + ligase	5.5	86	8.5	3931
pMS7 cut with <i>Bsr</i> FI + ligase	0	8.5	91.5	241
pAC2 cut with <i>Bsr</i> FI - ligase	9	75	15	168
pMS7 cut with <i>Bsr</i> FI - ligase	-	-	-	0
pAC2 uncut	10.5	80	9.5	1.038×10^7
pMS7 uncut	-	1	99	1.03×10^6

Figure 5.1

a) Negative picture of ethidium bromide stained agarose gel; *Pvu*II restriction digests of pAC2, pDIR1, pMS7 and pUC18.

b) Maps of the monomeric components of the plasmids analysed by restriction with *Pvu*II. pAC2 contains major proportions of the 893bp *Pvu*II fragment containing the 571bp palindrome and the 366bp *Pvu*II fragment containing the DIR. It also contains a minor proportion of the 322bp *Pvu*II fragment, characteristic of pUC18. pDIR1 contains equal proportions of the 366bp *Pvu*II DIR fragment and the 322bp *Pvu*II pUC18 fragment. pMS7 contains only the 893bp *Pvu*II palindrome-containing fragment. pUC18 contains only the 322bp *Pvu*II fragment (B = *Bsr*FI).

Figure 5.1

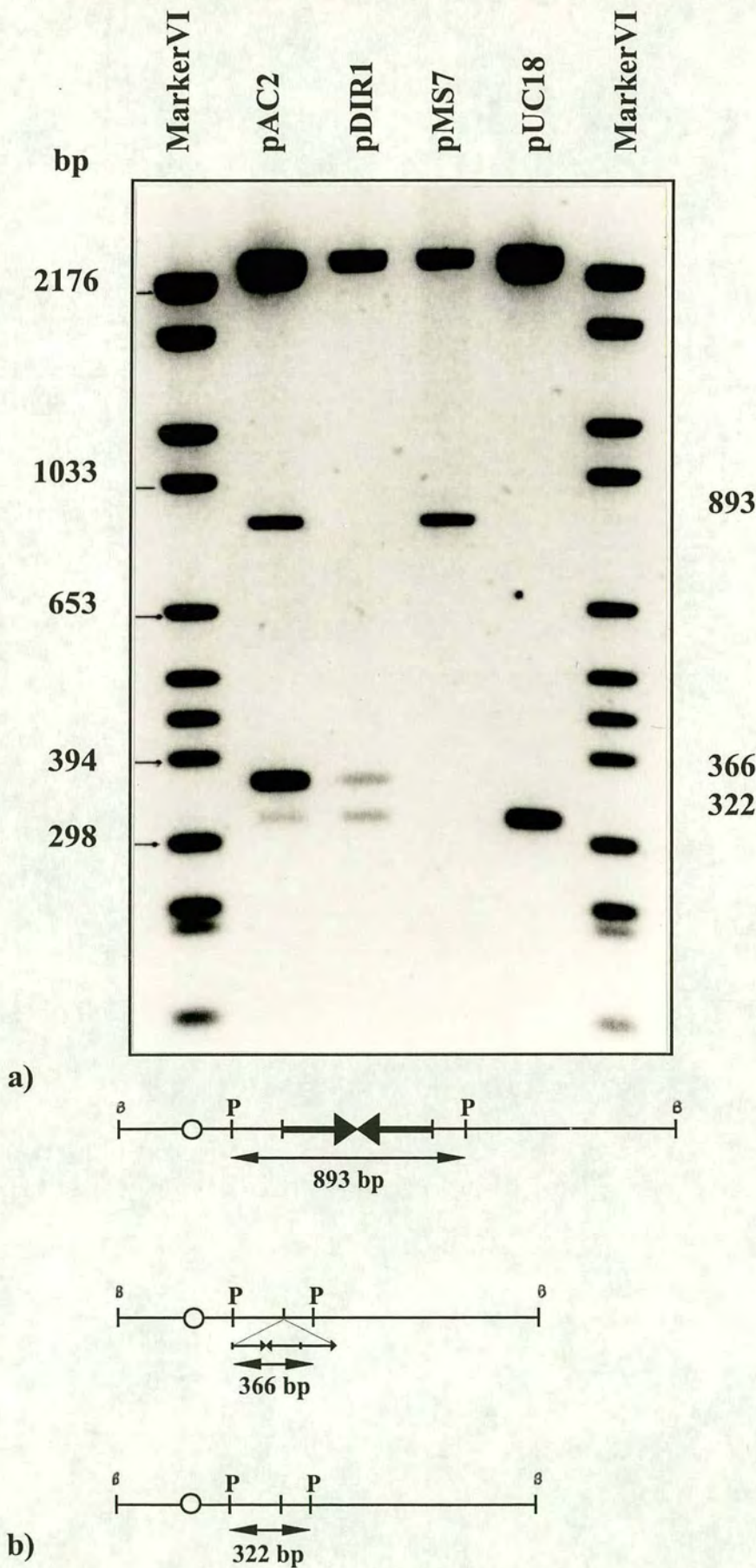


Figure 5.2

Autoradiograph of *Eco*RI single digests and *Eco*RI/*Bam*HI double digests of pAC2, pDIR1, pMS7 and pUC18.

Figure 5.2

Marker V

8-32

	EcoRI	EcoRI/BamHI
pAC2		
pDIR1		
pMS7		
pUC18		

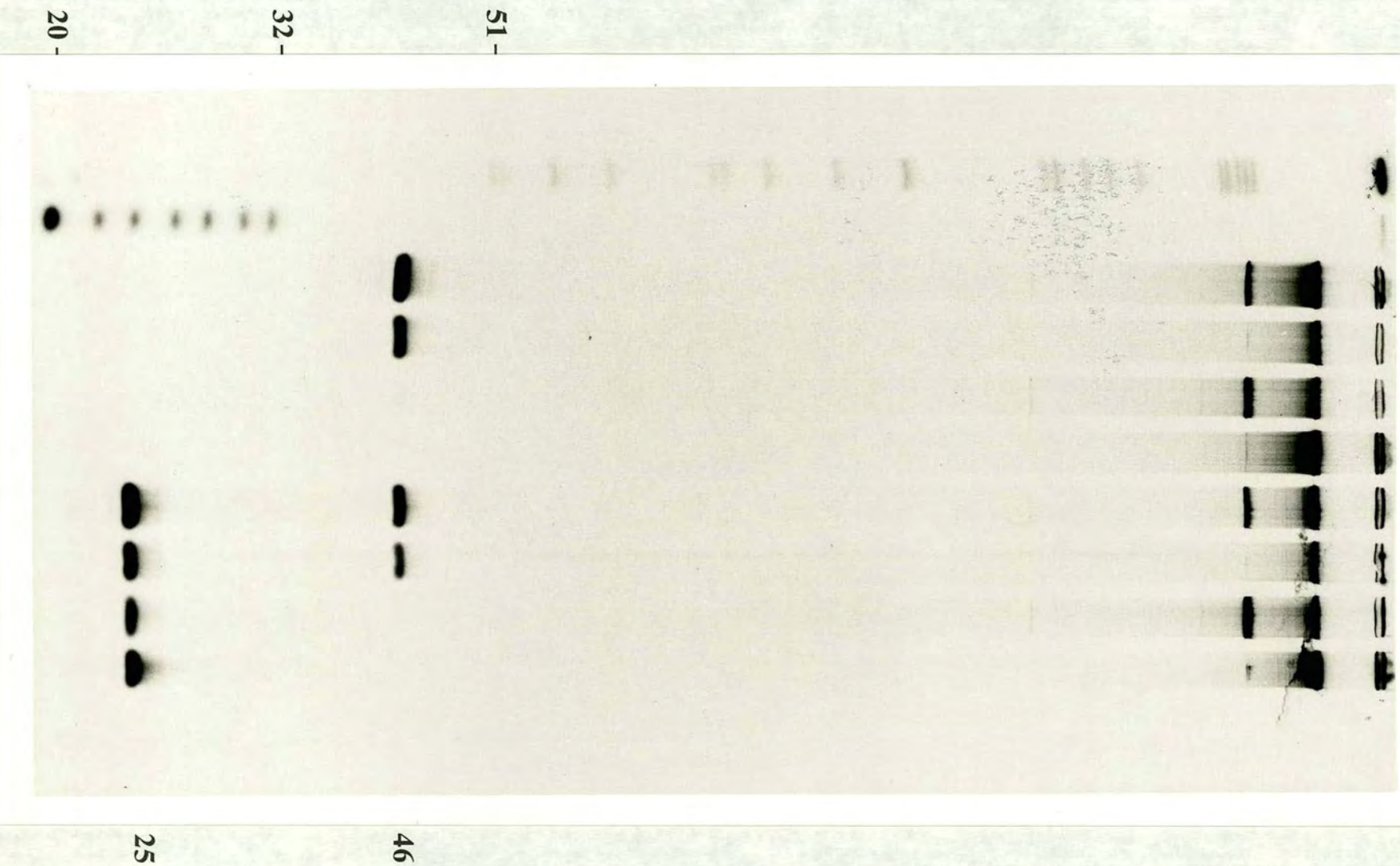


Figure 5.4

a) Autoradiograph of the DIR sequence. The data shown were obtained using primer 596L on pUC18 and on the homo-multimer substrate derived from pDIR1 after transformation of JM83. i) DIR sequence using the standard sequencing method. ii) pUC18 sequence using the standard method. iii) DIR sequence using the DMSO method to minimise pausing of Sequenase[®] at regions of secondary structure. The complete sequence of DIR could not be obtained on any individual gel but all ambiguities could be resolved using the DNA sequence obtained from the -40 primer (data not shown).

b) Sequence of pUC18 multicloning site and DIR: i) sequence of the multicloning site of pUC18, ii) sequence of the multicloning site in pDIR1, iii) diagrammatic representation of DIR.

Figure 5.4

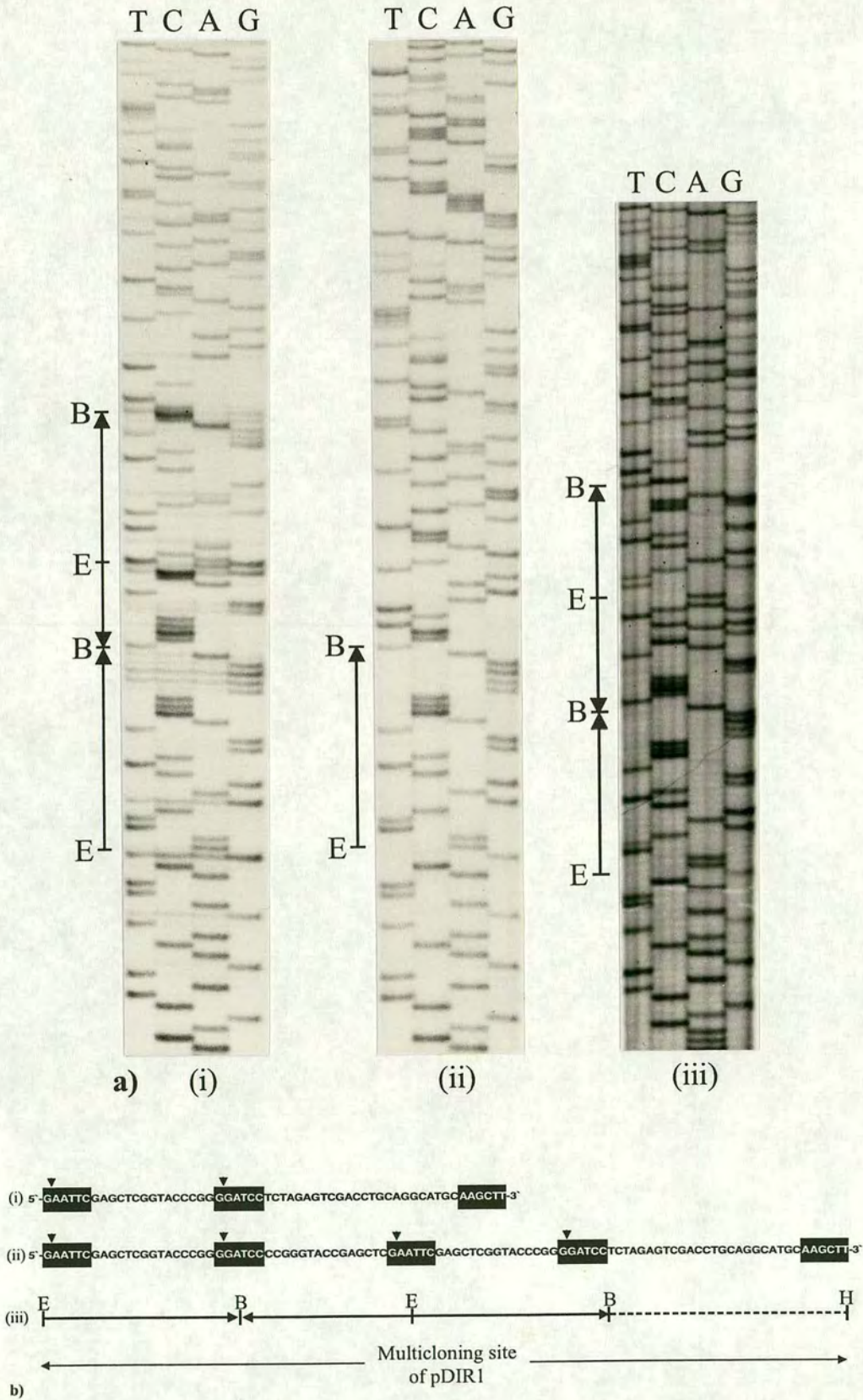
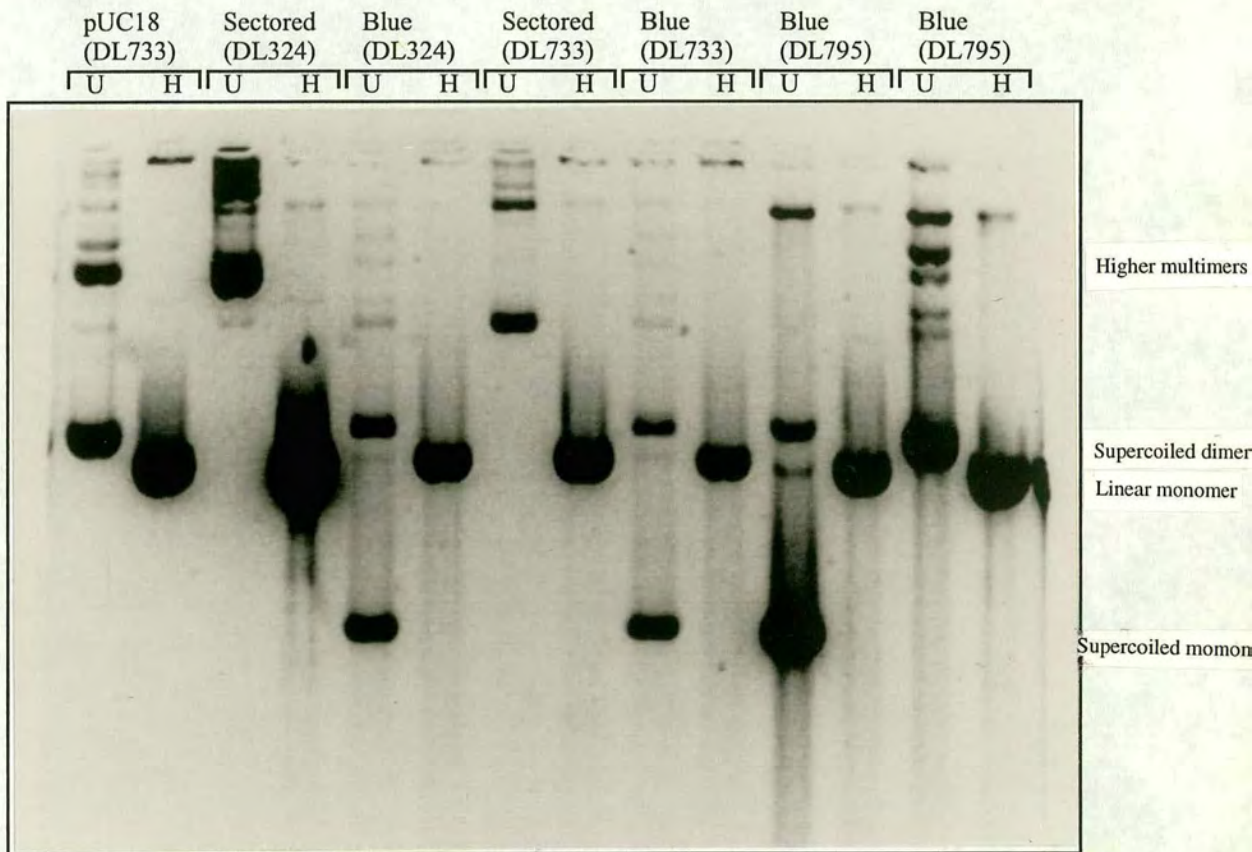


Figure 5.5

Uncut pDIR1 DNA and *Hind*III restriction digests of pDIR1 DNA isolated from different genetic backgrounds (DL324;DL733;DL795),(U = uncut; H= *Hind*III digest). Clones were Blue or had a sectored phenotype on AXI medium.



Conclusion

When pAC2 and pMS7 were monomerised, sectoring clones were discovered that did not contain the 571bp palindrome. One of these clones was analysed in detail and found to contain an inverted and directly duplicated region (DIR) in the polylinker adjacent to the site of insertion of the palindrome (see Fig. 5.4b for an overview of the structure). The palindrome itself was no longer present in the clones analysed. The same DIR structure was found to be present in pAC2 which is primarily a hetero-multimer composed of one monomeric subunit containing the palindrome and the other containing the DIR structure. pMS7 does not contain the DIR structure and is a homo-multimer, both subunits containing the palindrome.

As described previously, the most logical conclusion concerning the phenotype of pAC2, is that the sectored phenotype is due to instability of the 571bp palindrome. The isolation of 'DIR' provides an alternative explanation for the sectoring phenotype of pAC2. Only plasmids containing the DIR structure cause sectoring colonies, and it is possible to have sectoring clones which do not contain the palindrome. Therefore, one may conclude that the sectoring colony phenotype seen in pDIR1 and pAC2 is due to the presence of the DIR structure.

Since all these plasmids were maintained in Rec⁺ cells interconversion between various multimeric forms will be occurring to generate sub-populations of plasmids. The results shown in Table 5.1 are consistent with heterogeneous populations of plasmids in the DNA preparations of pAC2 and pMS7. They also show it is possible to isolate the sub-populations in multimeric populations of plasmids by forcing monomerisation. Additionally, monomerisation may actually stimulate rearrangement, this will be discussed in Chapter 7.

The DIR structure itself is unstable and it deletes out in a perfect manner to generate pUC18. It is not known if the sectoring phenotype observed when the DIR structure is present is due to its deletion, or if some other phenomenon is responsible for the sectored phenotype. This will be discussed in Chapter 7.

NOTE: At the conclusion of the work described in this chapter, pM*167 and pM*180 DNA, were restricted with *PvuII*, to check for the presence of DIR. pM*167 and pM*180, were found not to have the DIR mutation (data not shown).

CHAPTER 6:

The construction of cDR cloning vectors

Introduction

Cosmid vectors were first developed by Collins and Hohn (1978). Cosmids may be defined as plasmids containing the *cos* region of bacteriophage λ , an antibiotic resistance gene for selection and unique restriction sites for insertion of foreign DNA. Cosmid cloning utilises the natural mechanisms of lambda phage, to transform *E. coli*. For an overview of *in vitro* packaging of cosmid DNA see Chapter 1 of this thesis.

Deletion has been closely associated with cosmids and other similar large capacity cloning vectors. Collins and Hohn (1978) noticed the absence of palindromic structures and noted a large number of deletions among their isolates. There have since been many anecdotal reports of instability of human DNA inserts in genomic cosmid libraries (Chia *et al.*, 1982; Steinmetz *et al.*, 1982; Yelton *et al.*, 1985).

Deletion due to homologous recombination can be prevented by using *recA* strains (Kurnit, 1989) and a *recB recC sbcB recJ* host can prevent some illegitimate recombination events (Ishiura *et al.*, 1989; Ishiura *et al.*, 1990). However, there is still a requirement for a cloning system which reduces the current levels of illegitimate recombination in *E. coli*.

At the beginning of this work, the goal was to construct a deletion resistant cosmid. A 'mutant' plasmid called pMS7 which seemed to stabilise a 571bp palindrome had been isolated; the idea was to use the part of pMS7 containing the 'mutation' as part of the backbone for a new cosmid. Chapters 3 to 5 in this thesis describe how pMS7 was analysed; the results demonstrated that pMS7 does not contain a mutation. In this Chapter the construction of three new cosmids are described. The first cosmid constructed, cDRI, is referred to briefly, while the construction of the remaining cosmids, cDRII and cDRIII are reviewed in detail.

Figure 6.1

Overview of the construction of cDRI(i).

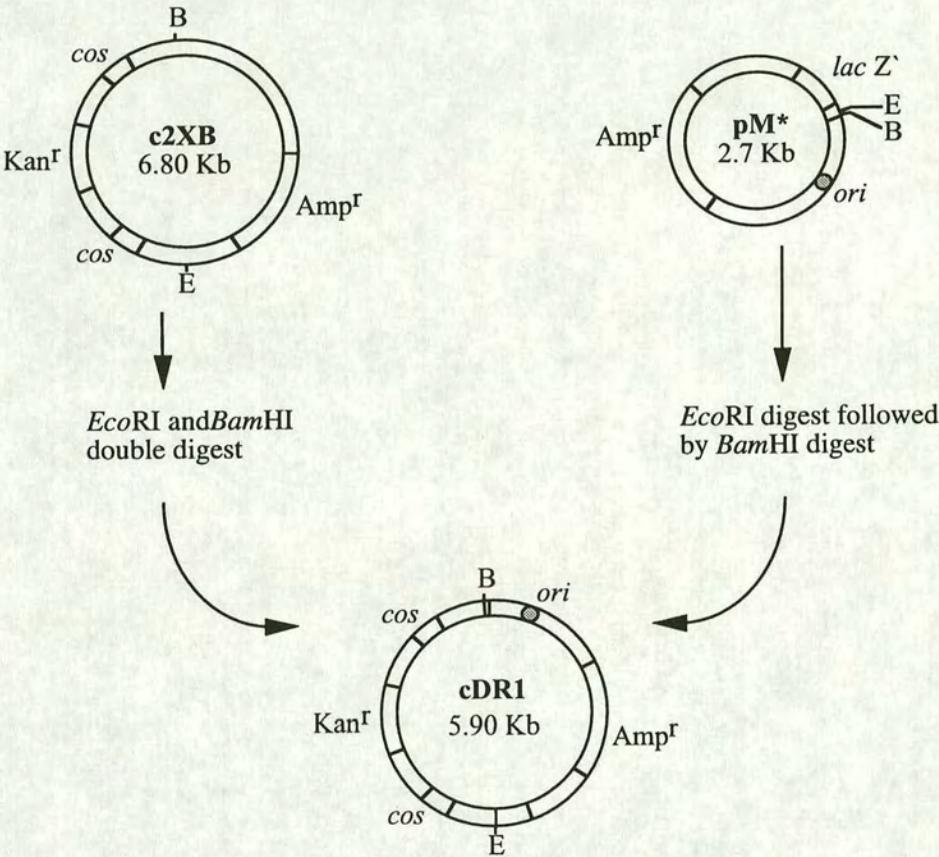
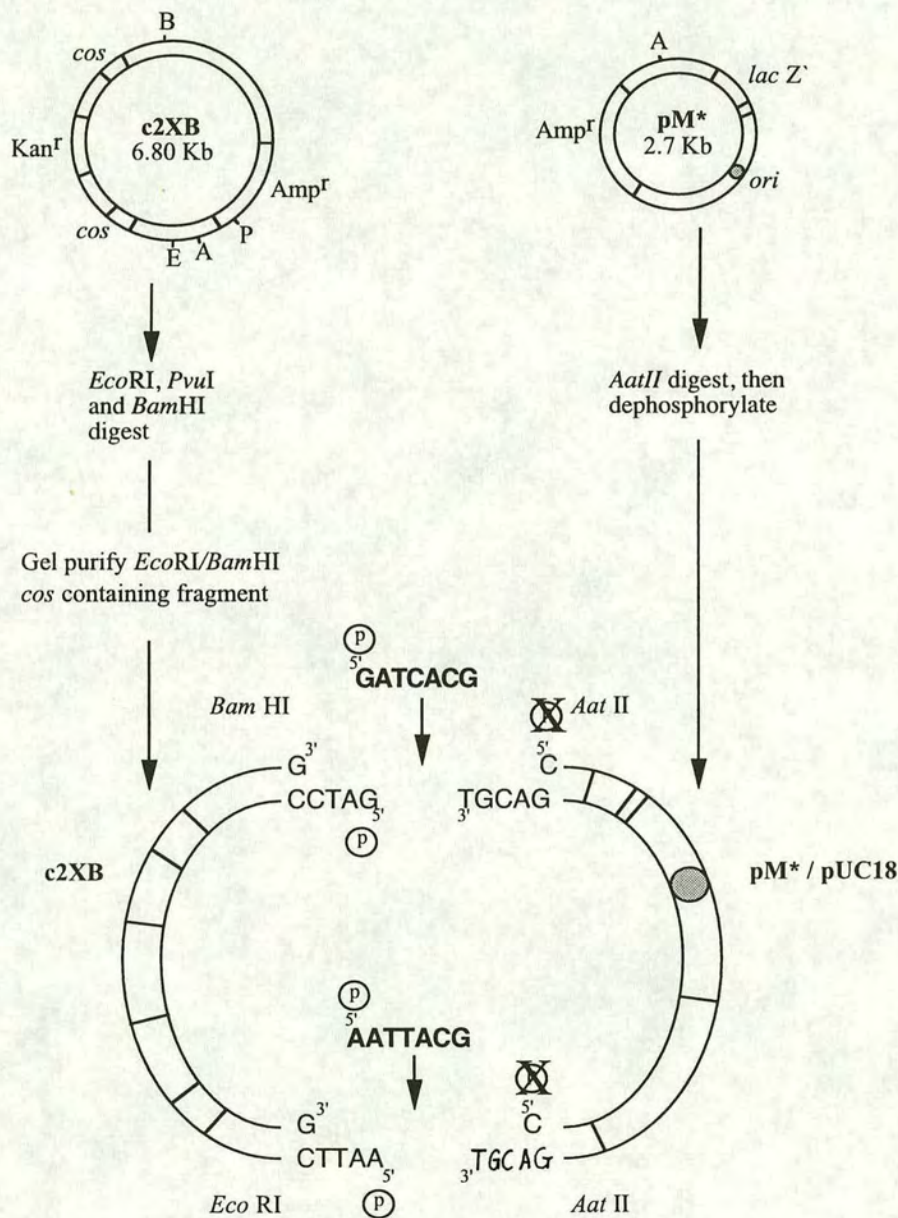


Figure 6.2

Overview of the construction of cDRI(ii).



The ligation mixture was recut with *Aat II*, to cut pM* and also screen the unwanted fragment in c2XB.

The construction of cDRI

The construction of cDRI involved the insertion of the *cos* containing *EcoRI* to *Bam*HI fragment from c2XB into pM*. c2XB is a double *cos* vector with a kanamycin resistance gene between the *cos* sites for selection and screening against small inserts (Bates & Swift, 1983).

The construction of cDRI proved difficult, and as a result two methods of construction were employed; these methods are summarised in Fig. 6.1 & 6.2. After many attempts cDRI was finally constructed using the method outlined in Fig. 6.1. When cDRI DNA was digested with *EcoRI* and analysed by restriction and gel electrophoresis it was found to be unstable. cDRI appeared to delete part of its backbone to give a smaller plasmid of about 2kb. The cosmid was maintained in a *recA* host (1046; Bates & Swift, 1983), so deletion was likely to be due to an illegitimate recombination event. Attempts were made to use cDRI for the construction of a cosmid library, but the work was abandoned when instability of the vector continued (some of the work on analysis of cDRI was done by Thorsten Allers).

The Construction of cDRII & cDRIII

Following the problems associated with cDRI, a different approach was applied in the construction of cDRII and cDRIII. These cosmids were engineered by inserting two oligos into pM*. The oligos contained *cosQ* and a number of restriction sites, to facilitate the insertion of two *cos* containing *Mbo*I fragments from lambda and the kanamycin resistance GenBlock from pUC4K. This construction required sequencing to check the orientation of a ligated fragment. The sequencing primers used may be seen in Fig. 6.3.

Figure 6.3

Primers for sequencing in both directions towards the *Aat*II site of pUC18.

(PRIMER1 :) 5`-AAAGGGCCTCGTGATACG-3`

(PRIMER2 :) 3`-GGCGAGTACTCTGTTATT-5`

A summary of the structure of *cos* DNA in λ , the sequences of the oligos, the construction of the cosmids and a scheme for use of the cosmids are summarised in Fig. 6.4a-i, and in the following text:

a) Summary of the structure of *cos* DNA in λ and the 431bp *Mbo*I λ fragment

As described previously in Chapter 1 of this work, the *cohesive-end site* (*cos*) comprises about 200bp and consists of several subunits. They include: i) *cosB*, which contains a series of protein binding sites ii) *cosN*, where the DNA is nicked specifically by terminase to generate the 12bp cohesive ends, and iii) *cosQ*, which contains the R4 element (originally classed as part of *cosB*). The 431bp *Mbo*I λ fragment used in this construction does not contain *cosQ*. However, *cosQ* is incorporated into the construction as part of the oligo sequences. The oligos also contain a number of restriction sites.

b) Isolation of the *cos* Containing fragment from Lambda

The 431bp *Mbo*I fragment from lambda contains *cosB* and *cosN* and was isolated in the following manner. Lambda DNA (wild type), was ligated to form a circular molecule. This was digested with *Pst*I, and the 14057bp fragment generated, was purified by gel electrophoresis. This fragment was ligated into pBR322 and cloned in *E. coli* JM83. Recombinant plasmid DNA was isolated, digested to completion with *Mbo*I and the 431bp fragment generated was purified by gel

electrophoresis, (recombinant plasmid DNA (pBR322 + the 14057bp *Pst*I fragment from λ), was provided by Ewa Okely).

c) Ligation of Oligo1 into pM*

Single stranded oligonucleotides obtained from Oswel DNA service were annealed to form double stranded Oligo1. pM* was digested with *Aat*II.

Oligo1 has *Aat*II-like ends, and was ligated into the *Aat*II site of pM*; this destroys the original *Aat*II site in the plasmid. An *Aat*II digest was employed to reduce background, before the DNA mixture was used to transform *E. coli* DL887 (*recA*). Recombinant plasmid DNA was isolated using a QIAGEN Plasmid Midi Kit. Oligo1 can be ligated in two orientations so, the recombinant DNA was sequenced to determine if in the correct orientation.

d) Ligation of *Mbo*I λ Fragment into Oligo1, in pM*

Oligo1 contains the R4 site (*cosQ*), and on the right side of R4 there is a *Bcl*I site. *Bcl*I is sensitive to Dam methylation, so recombinant plasmid DNA from the previous step was used to transform GM119 (*dam*). plasmid DNA was then isolated and digested with *Bcl*I which generates GATC overhangs; these 4bp overhangs are complementary to the overhangs generated by *Mbo*I. The purified 431bp fragment DNA from part b) was ligated into the *Bcl*I site in Oligo1; this destroys the *Bcl*I site. Recombinant plasmid DNA was isolated using a QIAGEN Plasmid Midi Kit. The *Mbo*I fragment can be ligated in two orientations so, the recombinant DNA was sequenced to determine if in the correct orientation.

e) Ligation of Oligo2 into Oligo1, in pM*

Single stranded oligonucleotides obtained from Oswel DNA service were annealed to form double stranded Oligo2. DNA from the previous step was digested with *Xho*I. Oligo2 has *Xho*I like ends; when Oligo2 is ligated into the *Xho*I site in Oligo1, the *Xho*I site is destroyed. An *Xho*I digest was used to reduce background, before the DNA mixture was used to

transform *E. coli* DL887 (*recA*). Recombinant plasmid DNA was isolated using a QIAGEN Plasmid Midi Kit. Oligo2 can be ligated in two orientations so, the recombinant DNA was sequenced to determine if in the correct orientation.

f) Ligation of *Mbo*I λ Fragment into Oligo2

Oligo2 contains the R4 site (*cosQ*), and on the right side of R4 there is a *Bcl*I site. *Bcl*I is sensitive to Dam methylation so recombinant plasmid DNA from the previous step was used to transform GM119 (*dam*). plasmid DNA was isolated and digested with *Bcl*I which generates GATC overhangs; these 4bp overhangs are complementary to the overhangs generated by *Mbo*I. The purified 431bp fragment DNA from part b) was ligated into the *Bcl*I site in Oligo2; this destroys the *Bcl*I site. Recombinant plasmid DNA was isolated using a QIAGEN Plasmid Midi Kit. The *Mbo*I fragment may be ligated in two orientations so, the recombinant DNA was sequenced to determine if in the correct orientation.

g) Isolation of the Kanamycin GenBlock

pUC4K contains the Kanamycin Resistance GenBlock and is available commercially from Pharmacia Biotech. The GenBlock is approximately 1.3kb long and contains the drug resistance marker from the transposon Tn903; it was isolated from pUC4K. pUC4K DNA was digested with *Bam*HI and the 1264bp fragment containing the kanamycin resistance gene was purified by agarose gel electrophoresis.

h) Ligation of Kanamycin GenBlock into *Bgl*II site in Oligo1

DNA from f) was digested with *Bgl*II and the ends were dephosphorylated. The Kanamycin GenBlock has GATC overhangs generated from a *Bam*HI digest. The GenBlock was ligated into the *Bgl*II site in Oligo1; this destroys the *Bgl*II site. A *Bgl*II digest was employed to reduce background, before the DNA mixture was used to transform *E. coli* DL887 (*recA*). Recombinant plasmid DNA was isolated using a QIAGEN Plasmid Midi Kit. The Kanamycin GenBlock may be ligated in two

orientations so recombinant DNA was digested with *Xho*I, *Bsr*FI and *Xho*I/*Bsr*FI (*Xho*I cuts once in the kanamycin resistance gene, while *Bsr*FI cuts twice, once in the kanamycin resistance gene and once in the ampicillin resistance gene). The fragments generated were analysed by agarose gel electrophoresis to determine which orientation the GenBlock was inserted in each cosmid.

cDRII contains the kanamycin Gene in one orientation and cDRIII contains the kanamycin Gene in the opposite orientation. The full sequences of cDRII and cDRIII are in Appendix I of this work.

i) Scheme of Use for cDRII & cDRIII

cDRII and cDRIII are double *cos* vectors; this eliminated the need to prepare two separate cosmid arms, which is a long and tedious process. They have a kanamycin resistance gene between the *cos* sequences for selection and screening against small inserts and are based on pUC18. Sites available for insertion of foreign DNA include, *Eco*RI, *Bam*HI and any site in the polylinker region of *lacZ'*, not present elsewhere in the vector. The cosmids are 4.9kb and therefore have insert limits of approximately 34-47kb. Their use in the construction of a cosmid library is summarised in Fig. 6.4i.

The cloning efficiency of the cDR cosmids

In order to check the cloning efficiency of the new cosmid vectors, human genomic DNA was ligated into the *Bam*HI site of cDRII and cDRIII, and into SuperCos as a positive control. The recombinant DNA was packaged using Gigapack® II XL Packaging Extracts. After packaging, 500µl SM buffer were added to the packaging reaction, then 200µl of the diluted packaged mixture were added to 250µl of the host strain (DL795). Various dilution's of the mixture were then plated on L-agar supplemented with ampicillin (100µg/ml).

Results

The first cloning experiments with cDRII, cDRIII and SuperCos yielded very low numbers of clones (~4/plate). Extrachromosomal DNA was isolated and uncut DNA was analysed by agarose gel electrophoresis. As Fig. 6.5 shows only three of the eight clones chosen contained high molecular weight recombinant cosmid DNA.

Further cloning experiments using fresh genomic DNA were performed. Both cDRII and SuperCos yielded larger numbers of clones. However, cDRII seemed more unstable than SuperCos. Most of the SuperCos clones analysed, contained high molecular weight recombinant cosmid DNA. cDRII clones consistently had lower number of high molecular weight recombinant cosmid DNA, and many of the isolates analysed, had undergone deletion.

Note: The text sections a-i, pp102-105, correspond to the sections a-i, in Fig. 6.4, pp107-115. The sections in brackets are repeated from previous page(s).

Figure 6.4
 Overview of the construction of cDRII and cDRIII.

Figure 6.4a)

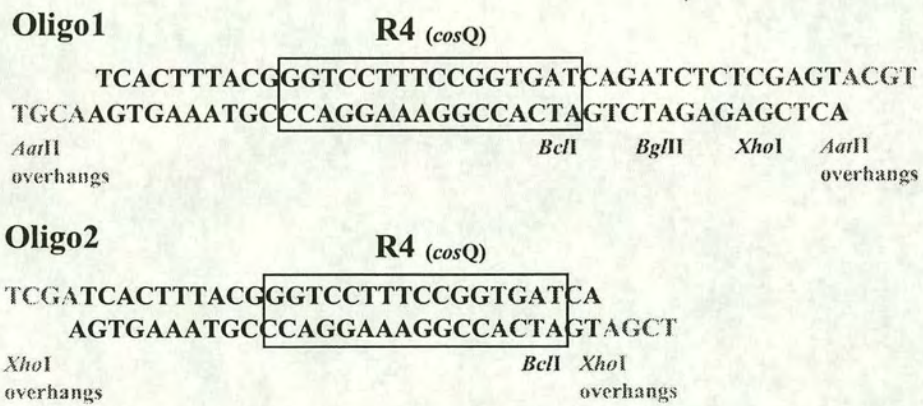
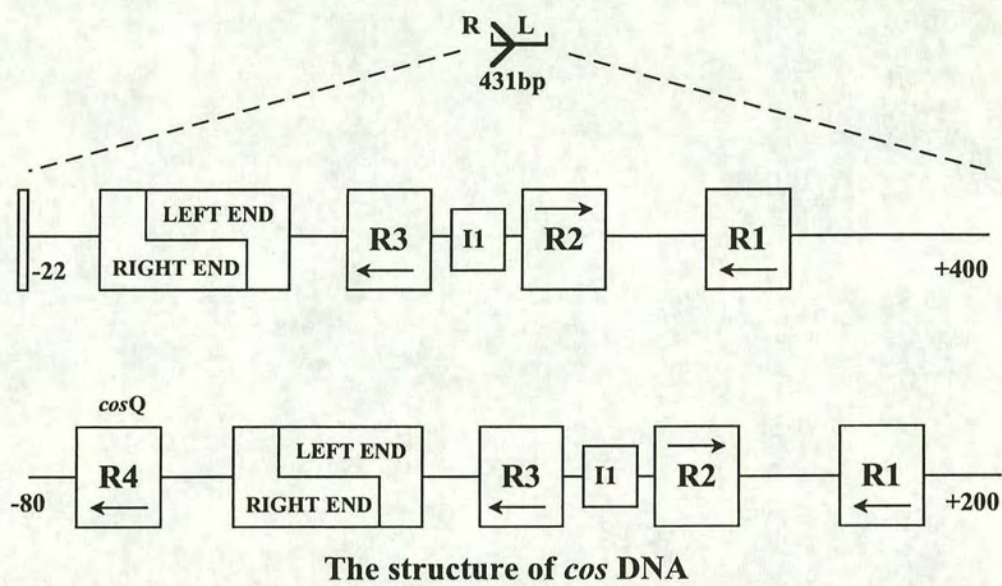


Figure 6.4b)

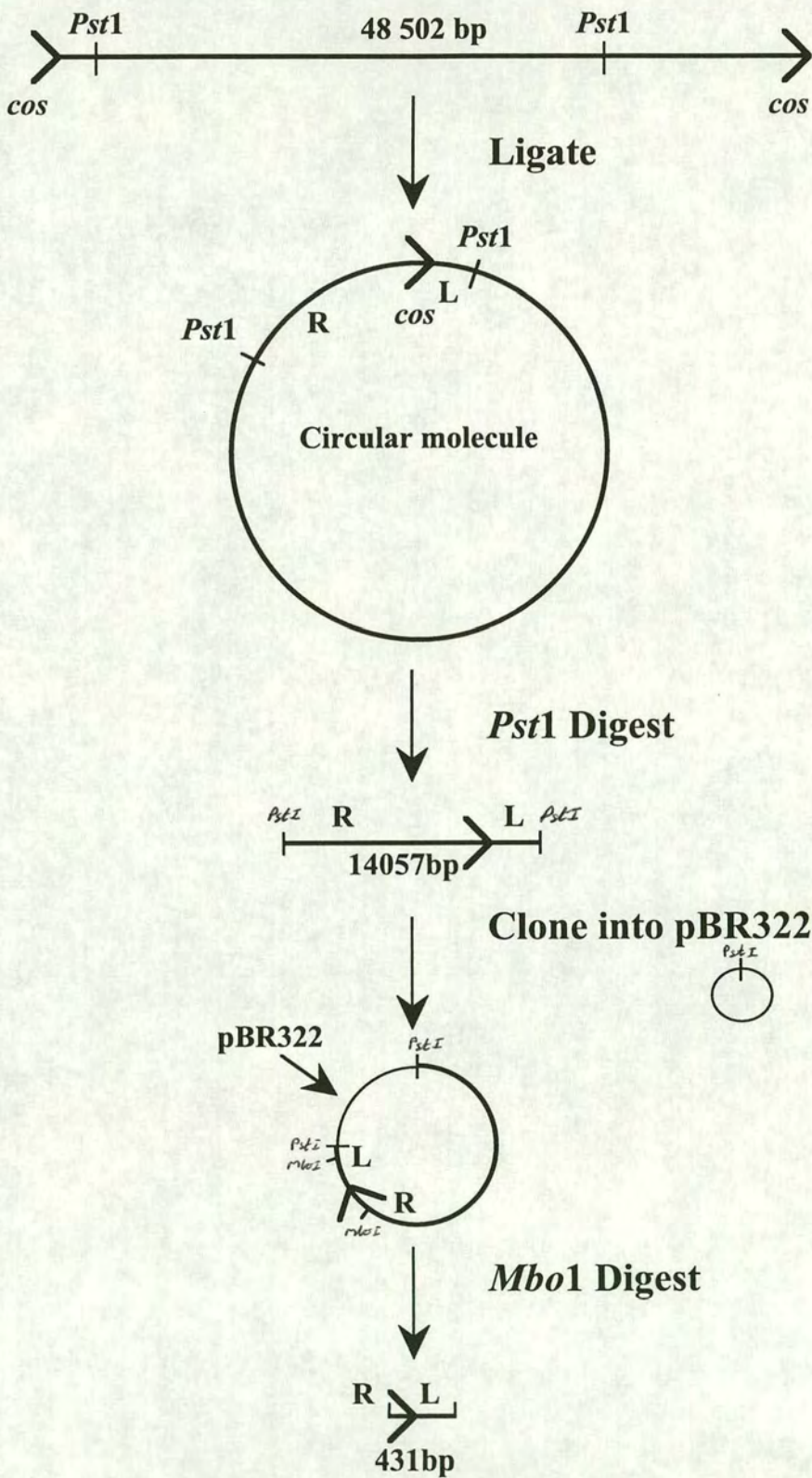


Figure 6.4c)

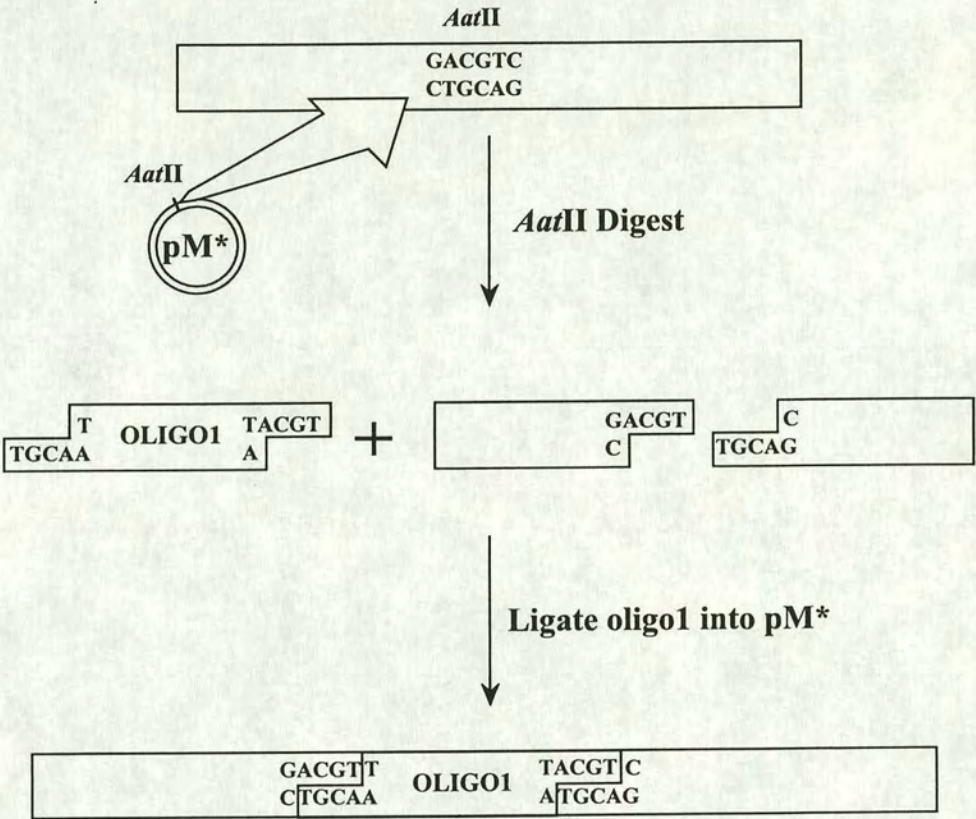


Figure 6.4d)

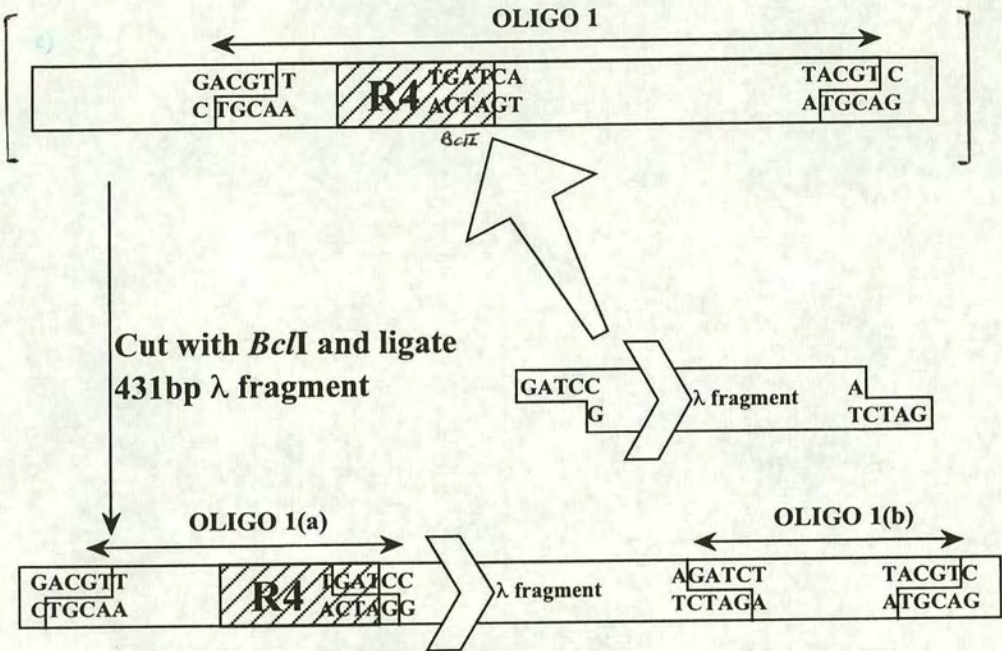


Figure 6.4e)

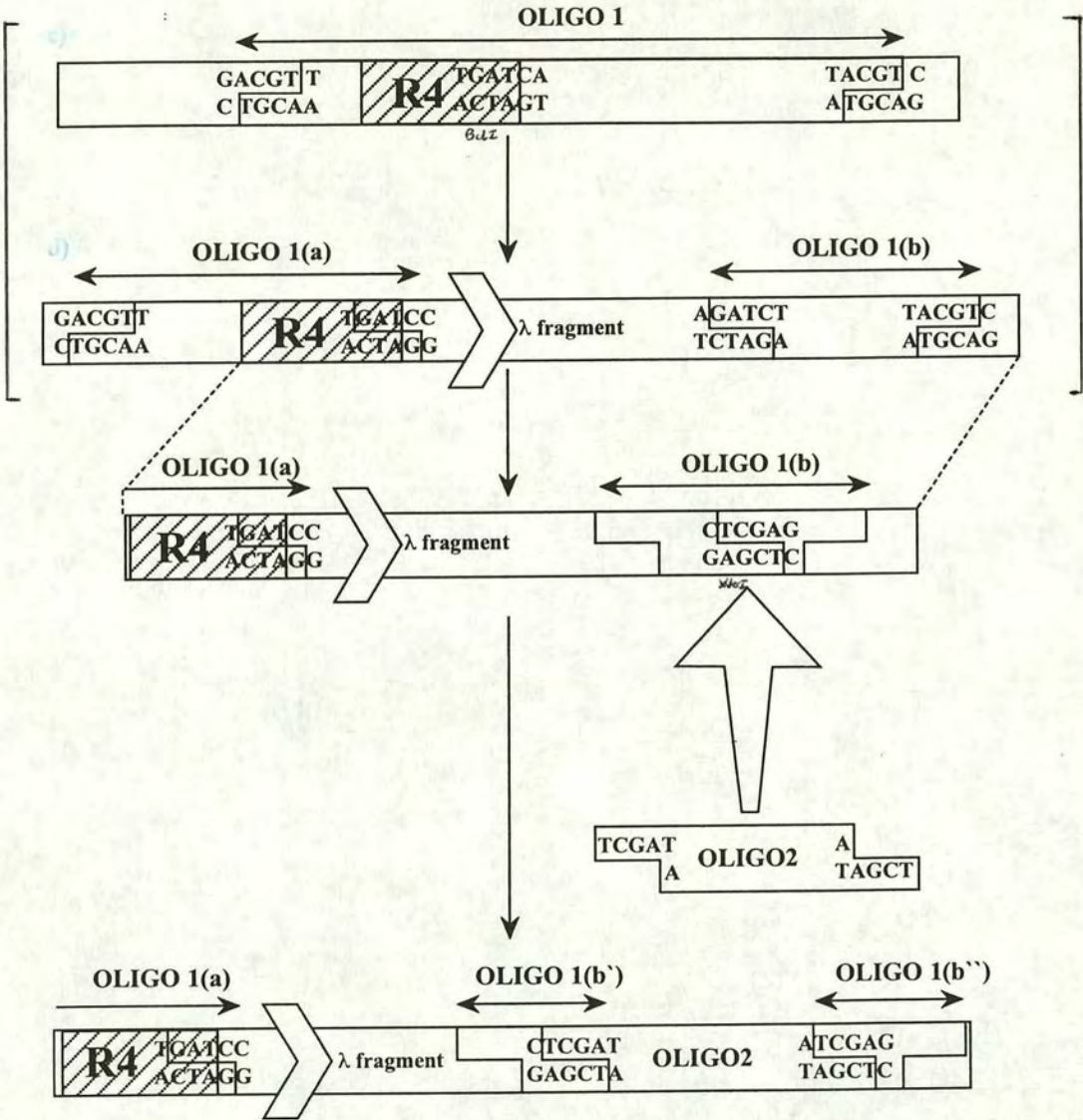


Figure 6.4f)

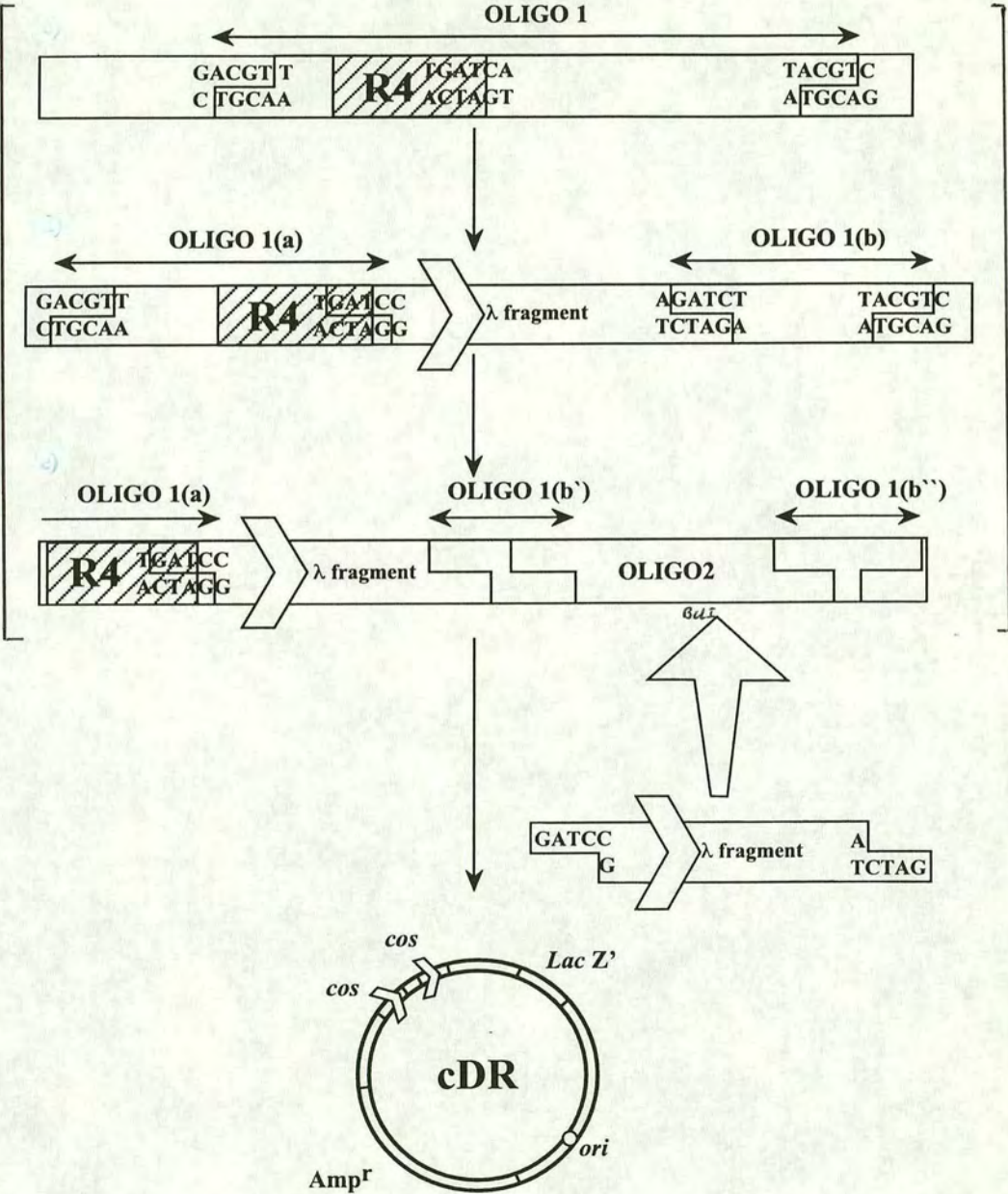


Figure 6.4g)/h)

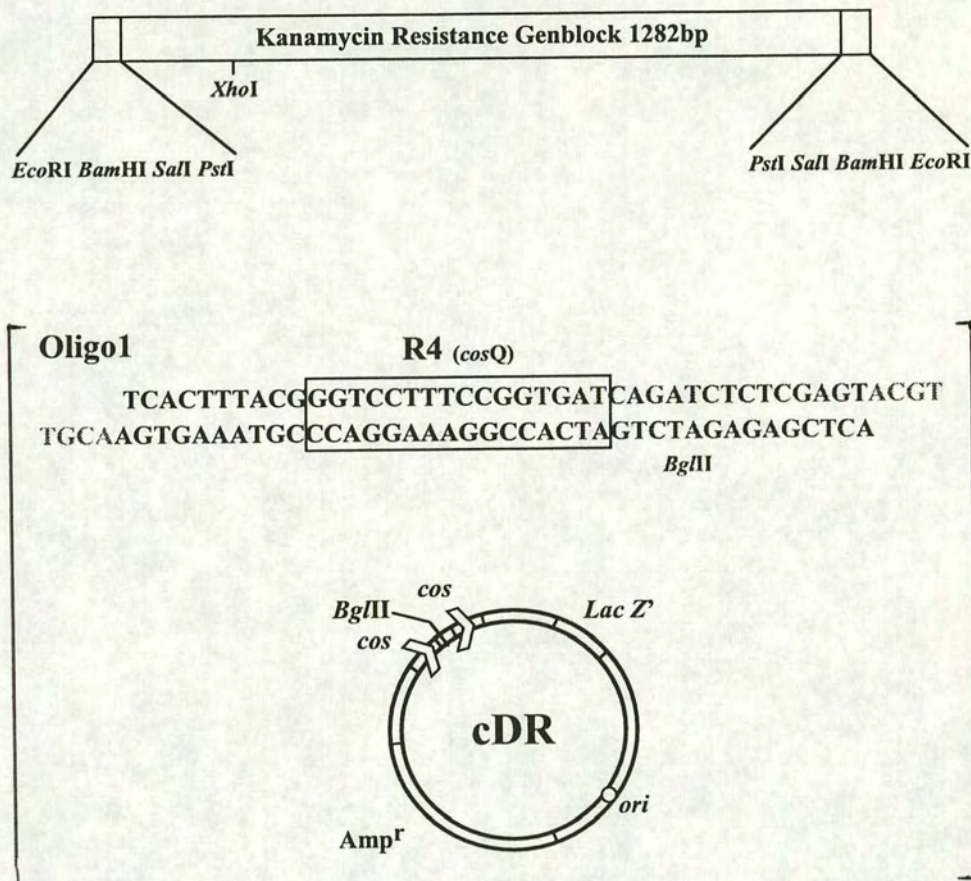


Figure 6.4h)

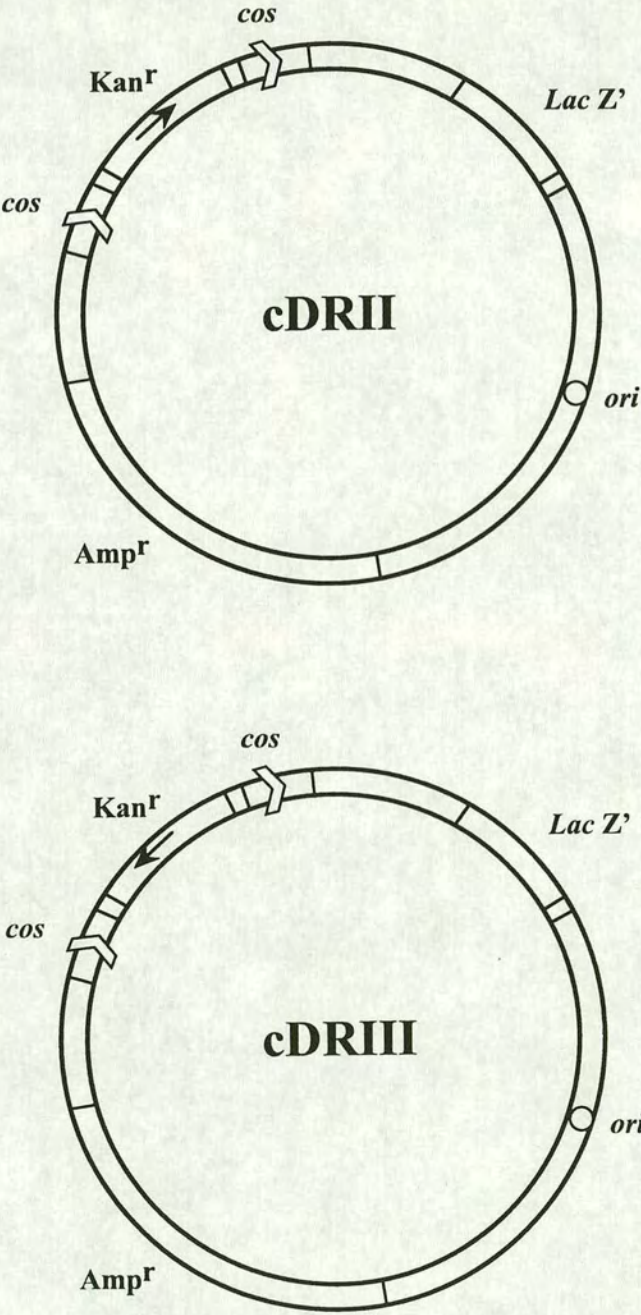


Figure 6.4i)

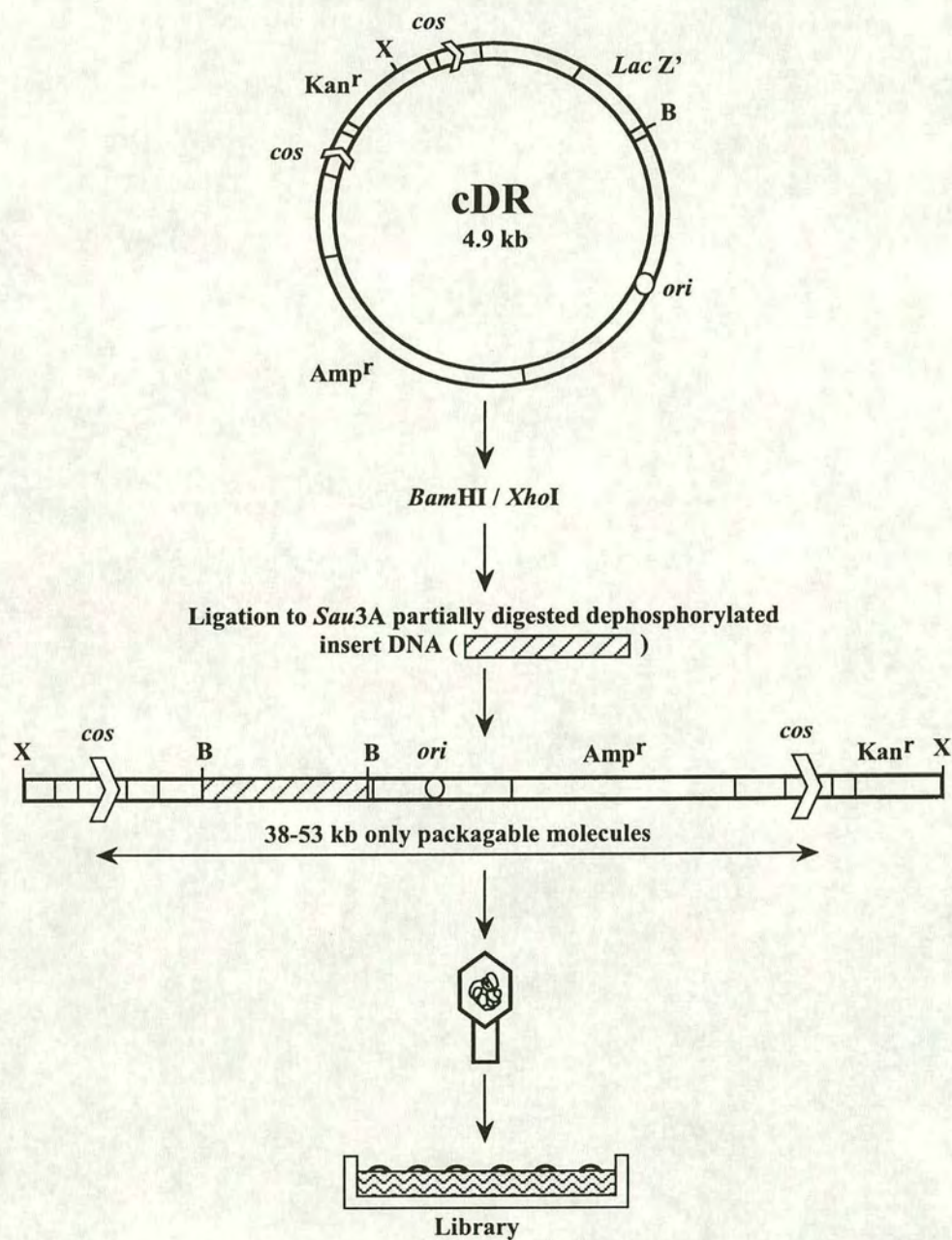
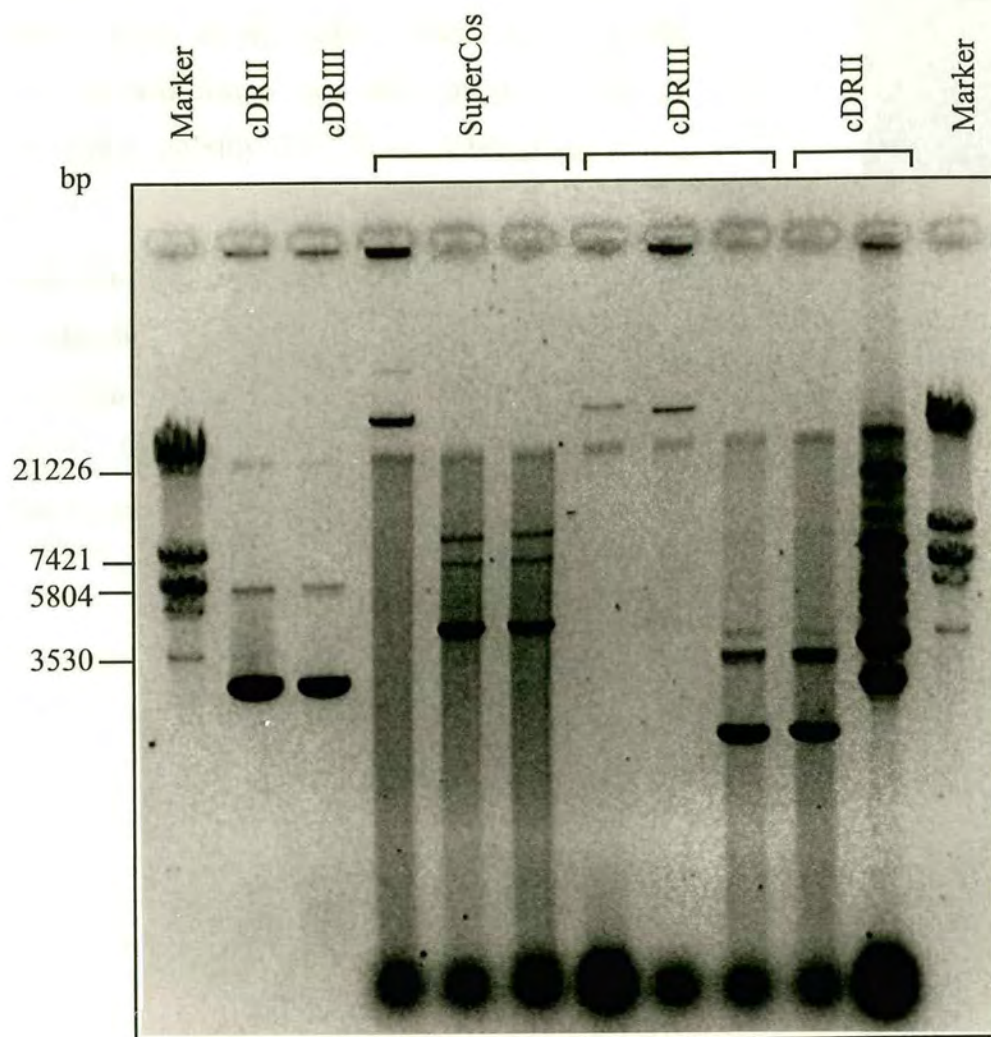


Figure 6.5

Negative picture of ethidium bromide stained agarose gel: uncut cDRII, III and SuperCos recombinant DNA.



Conclusion

cDRII and cDRIII have been constructed. They both may be used to clone large fragments of eukaryotic DNA; however, they are both subject to instability problems. This could be due to many reasons but, the most likely explanation for instability is, the high copy number origin of replication present in cDRII and cDRIII. Many other studies have reported that a low copy number origin of replication bestows an increase in cloning stability (Kim *et al.*, 1992; Jiang *et al.*, 1987; Little & Cross, 1985; Nakano *et al.*, 1995; Speek *et al.*, 1988).

The original purpose of constructing these vectors was, to construct a deletion resistant cosmid using the 'mutant' plasmid pMS7 which seemed to stabilise a 571bp palindrome. An alternative explanation for the behaviour of pMS7 has been described in Chapters 3-5 of this thesis. This means that the backbone of cDRII and cDRIII is likely to be pUC18 and therefore, is unlikely to contain a "stabilising mutation".

CHAPTER 7:

DISCUSSION

CHAPTER 7:

DISCUSSION

Summary

At the onset of this work the aim was to construct a deletion resistant cosmid using pM*, as part of the backbone. pM* was isolated from pMS7, by removing the 571bp palindrome with an *EcoRI* digest. pMS7 appeared to contain a 'mutation' which stabilised the propagation of the 571bp palindrome. The concept was, if a 'mutation' could stabilise a palindrome, then the 'mutation' could perhaps stabilise other unstable sequences. The results presented in Chapter 5 provided an alternative explanation for the phenotype of clones containing pMS7 when plated on AXI media. It was determined that pAC2 contained an unusual mutation, consisting of an inverted and directly duplicated region, (DIR).

This thesis describes the fine structure analysis of pAC2 and its derivatives; and the construction of two new cosmids using pM* as part of the backbone.

In Chapter 3, SSCP analysis, and a number of fragment swap experiments are described. The purpose of these experiments was to determine if there was a difference between pUC18 and pM*. No difference was observed between fragments of pUC18 and pM*, when run on SSCP gels. However, the fragment swap experiments did suggest a difference between pAC2 and pMS7; this was subsequently found to be due to the presence of DIR.

In Chapter 4, the cloning of four different palindromes, one with 10bp of asymmetry at its centre, and three perfect palindromes, are described. The purpose of these experiments was to demonstrate that these palindromes could be cloned in pM*, but not in pUC18. This would provide evidence for the presence of a 'mutation' in pM* and demonstrate its possible usefulness for the cloning of other unstable sequences. It was possible to clone all four palindromes in pM* but, not in pUC18. The results obtained provide evidence for differences between pUC18 and pM*. However, I invested much more time in attempting to ligate the palindromes into pM*. I anticipate that it might be possible to clone the four palindromes in pUC18, if as much effort was invested.

The copy number of pM*167 and pM*180 were also investigated. The results indicated a low copy number for each of the plasmids. Inviability and/or instability problems are known to be associated with the cloning of long DNA palindromes (reviewed by Leach, 1994); this provides an explanation for the observed low copy number of the plasmids. pM*180 appears to have a higher copy number than pM*167; this result confirms that asymmetry within a palindromic sequence influences the relative viability, of a plasmid containing a palindrome.

The results presented in Chapter 5 revealed that pAC2 is primarily a hetero-multimer composed of one monomeric subunit containing the 571bp palindrome and the other containing a novel mutation. This mutation was found to be composed of a direct and inverted repeat (DIR), of the *EcoRI* to *BamHI* site, in the polylinker region. pMS7 is primarily a homo-multimer and *E. coli* containing pMS7 is likely to represent, the normal situation for the propagation of the 571bp palindrome in pUC18. The sectoring phenotype observed when pAC2 clones are plated on AXI media is most likely due to the presence of the DIR structure. However, the direct mechanism of the production of pAC2 sectoring colonies is unknown.

Chapter 6 describes the construction of three new cosmid vectors. The first, cDRI, was constructed by ligating pM* to the *cos* containing fragment of c2XB. cDRI was found to be unstable so, further work on cDRI was discontinued. cDRII and cDRIII were constructed by ligating various fragments into pM*. These vectors have been used successfully as cosmid vectors. However, the evidence presented in previous chapters shows that pM* is unlikely to contain a mutation which stabilises palindromic sequences. Therefore, these cosmids are not likely to be deletion resistant. The cDR cosmids seem to suffer more instability problems than SuperCos when used to clone human genomic DNA. The most likely reason for this is the high copy number of pUC18, as most cosmid constructions report a low to medium copy number to be favourable.

Discussion

The results presented in this thesis describe the progressive search for an answer to the puzzle concerning, the differences in phenotype of clones containing pAC2 and pMS7 when plated on AXI media. The initial hypothesis regarding the sectoring phenotype of pAC2 was that it might be due to the 571bp palindrome deleting out from the *EcoRI* site in pUC18. It followed, that pMS7 might have contained a mutation which stabilised the 571bp palindrome. The discovery of the DIR mutation in pAC2 and pDIR1 provided an alternative conclusion. The sectoring “pAC2 like” phenotype observed was due to the presence of the DIR mutation. However, the mechanism for how the DIR mutation exerts this phenotype remains to be determined. No evidence for the presence of a mutation in pMS7 which stabilises palindromes was found.

In the following sections I wish to discuss, a model for the formation of the DIR mutation, and the future development of cDRII and cDRIII.

Model for the Formation of DIR

The DIR mutation was first discovered and analysed in pDIR1, a derivative of pAC2 that sectoried like pAC2, but did not contain the 571bp palindrome. The DIR structure was subsequently found in pAC2 also. Fine structure analysis of DIR revealed, it consists of a 42bp insert, which is both an inverted and direct repeat of the *EcoRI* to *BamHI* site, in the polylinker region. The structure of the DIR mutation is most easily explained if the 571bp palindrome facilitates strand-slippage at the replication fork from the leading to the lagging strand (as summarised in Fig. 7.1).

The first step, in the model for the formation of DIR, is the rare extrusion of the 571bp palindrome, to form a stable cruciform structure by intra-strand base-pairing. Once formed the secondary structure causes the DNA polymerase to pause. This facilitates replication across the fork from the leading to the lagging strand. A new palindromic strand is formed which can fold back to form a hairpin. leading

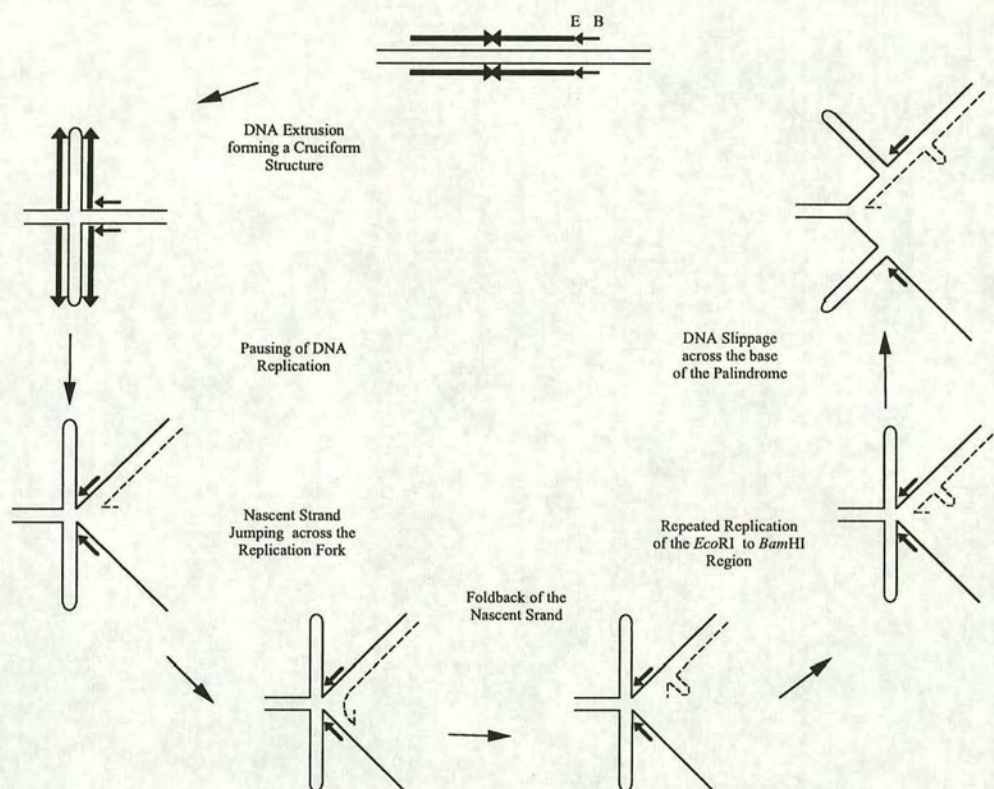
strand synthesis can then resume by re-copying the already duplicated region. A further strand-slippage event across the base of the extruded 571bp palindrome leads to the loss of the original palindrome. This event uses the *EcoRI* sites as a direct repeats and results in the loss of the original palindrome.

The model presented may only represent part of the mechanism for the formation of the DIR mutation. The complete model may be similar to the model proposed by Bi and Liu (1996; see Fig 1.3 in Chapter 1). In their model, replication slippage is associated with dimerisation, (a similar hypothesis is reviewed by Kuzminov, 1996). Further work would have to be completed, to establish such a link.

The mechanism for how the DIR mutation exerts the pAC2 phenotype remains to be determined. Two possibilities are likely: deletion of the DIR mediated by a further strand-slippage event, or interference of transcription of *lacZ* by DIR.

Figure 7.1

Model for the formation of DIR. The 571bp palindrome is represented by long bold arrows; the *EcoRI* -*Bam*HI fragment that is duplicated in DIR is represented by a short arrow; and the nascent leading strand is represented by a dashed line.



The future Development of cDRII and cDRIII

In their present form cDRII and cDRIII, seem to suffer more instability problems than SuperCos when used to clone human genomic DNA. The most likely reason being their high copy number as most cosmid constructions report a low to medium copy number to be favourable (see Chapter 1). Future developments of the cDR vectors would include the introduction of a medium to low copy number origin of replication and possibly, other useful features already present in other cosmids as previously described in Section 2 of Chapter 1.

BIBLIOGRAPHY

Ahmed, A. (1987). A simple and rapid procedure for sequencing long (40-kb) DNA fragments. *Gene* **61**, 363-372.

Ahmed, A. (1989). A vector for sequencing long (40-kb) DNA fragments. *Gene* **75**, 315-321.

Ahmed, A. (1994). Double-origin vectors for isolating bi-directional deletions useful in DNA sequence analysis. *Gene* **141**, 71-73.

Allers, T. (1993). Detection of Cruciform DNA *in Vivo*. Ph.D., University of Edinburgh.

Allgood, N. D. & Silhavy, T. J. (1988). Illegitimate recombination in bacteria. In *Genetic Recombination* (Kucherlapati, R. & Smith, G. R., eds.), pp. 309-330. American Society for Microbiology, Washington, DC.

Anderson, R. P. & Roth, J. R. (1977). Tandem genetic duplications in phage and bacteria. *Annu. Rev. Microbiol.* **31**, 473-505.

Arber, W. (1984). Natural Mechanisms of Microbial Evolution. In *Genetic Manipulation: Impact on Man and Society* (Arber, W., Illmensee, K., Peacock, W. J. & Starlinger, P., eds.), pp. 1-14. 3rd Cogene Symposium. Cambridge, Cambridge.

Bagdasarian, M., Lurz, R., Rückert, B., Franklin, F. C. H., Bagdasarian, M. M., Frey, J. & Timmis, K. N. (1981). Specific-purpose plasmid cloning vectorsII. Broad host range, high copy number, RSF1010-derived vectors, and a host-vector system for gene cloning in *Pseudomonas*. *Gene* **16**, 237-247.

Baldini, A., Ross, M., Nizetic, D., Vatcheva, R., Lindsay, E., A Lehrach H. & Siniscalco M. (1992). Chromosomal assignment of human YAC clones by

fluorescence *in situ* hybridization: use of single-yeast-colony PCR and multiple labelling. *Genomics* **14**, 181-184.

Bates, P. F. & Swift, R. A. (1983). Double *cos* site vectors: simplified cosmid cloning. *Gene* **26**, 137-146.

Bates, G. P., Valdes, J. Hummerich, H. Baxendale, S. Le Paslier, D. L. Monaco, A. P. Tagle, D. MacDonald, M. E. Altherr, M., Ross M., *et al.* (1992). Characterization of a yeast artificial chromosome contig spanning the Huntington's disease gene candidate region. *Nat. Genet.* **1**, 180-187.

Baumer, A., Zhang, C., Linnane, W. & Nagley, P. (1994). Age-related Human mtDNA deletions: A Heterogeneous set of deletions arising at a single pair of directly repeated sequences. *Am. J. Hum. Genet.* **54**, 618-630.

Bedinger, P., Munn, M. & Alberts, B. M. (1989). Sequence-specific pausing during *in vitro* DNA replication on double-stranded DNA templates. *J. Biol. Chem.* **264**, 16880-16886.

Bender, W., Spierer, P. & Hogness, D. S. (1983). Chromosome walking and jumping to isolate DNA from the Ace and rosy loci and the Bithorax complex in *Drosophila melanogaster*. *J. Mol. Biol.* **168**, 17-33.

Berg, D. E. & Howe, M. M., Eds. (1989). *Mobile DNA*. Washington D. C.: American Society for Microbiology.

Bi, X. & Liu, L. F. (1994). *recA*-independent and *recA*-dependent intramolecular plasmid recombination. *J. Mol. Biol.* **235**, 414-423.

Bi, X., Lyu, Y. L. & Liu, L. F. (1995). Specific stimulation of *recA*-independent plasmid recombination by a DNA sequence at a distance. *J. Mol. Biol.* **247**, 890-902.

Bi, X. & Liu, L. F. (1996). A replication model for DNA recombination between direct repeats. *J. Mol. Biol.* **256**, 849-858.

Bi, X. & Liu, L. F. (1996). DNA rearrangement mediated by inverted repeats. *Proc. Natl. Acad. Sci. USA* **93**, 819-823.

Bichara, M., Schumacher, S. & Fuchs, R. P. P. (1995). Genetic instability within monotonous runs of CpG sequences in *Escherichia coli*. *Genetics* **140**, 897-907.

Bierne, H. & Michel, B. (1994). When replication forks stop. *Mol. Microbiol.* **13**, 17-23.

Bijovet, J. F., van der Zanden, A. L., Goosen, N., Brouwer, J. & van de Putte, P. (1991). DNA insertions in the 'silent' regions of the 2 microns plasmid of *Saccharomyces cerevisiae* influence plasmid stability. *Yeast* **7**, 347-356.

Birnboim, H. C. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**, 1513-1523.

Blake, C. (1996). Illegitimate recombination in *Escherichia coli*. Ph.D., University of Edinburgh.

Boccard, F. & Prentki, P. (1993). Specific interaction of IHF with RIBs, a class of bacterial repetitive elements located at the 3' end of transcription units. *EMBO J.* **12**, 5019-5027.

Bouffler, S., Silver, A. & Cox, R. (1993). The role of DNA repeats and associated secondary structures in genomic instability and neoplasia. *BioEssays* **15**, 409-412.

Breter, H. J., Knoop, M. T. & Kirchen, H. (1987). The mapping of chromosomes in *Saccharomyces cerevisiae*: I. A cosmid vector designed to establish, by cloning into

cdc-mutants, numerous start loci for chromosome walking in the yeast genome. *Gene* **53**, 181-190.

Burke, D. T., Georges, G. F. & Olson, M. V. (1987). Cloning of large segments of exogenous DNA into yeast by means of Artificial Chromosome Vectors. *Science* **236**, 809-812.

Campbell, A. (1979). Natural modes of Genetic exchange. In *Recombinant DNA and Genetic Experimentation* (Morgan, J. & Whelan, W. J., eds.), pp. 21-29. Pergamon, Oxford.

Campbell, D. A. (1989). c2X75, a derivative of the cosmid vector c2XB. *Nucleic Acids Res.* **17**, 458.

Canning, S. & Dryja, T. P. (1989). Short direct repeats at the breakpoints of deletions of the retinoblastoma gene. *Proc. Natl. Acad. Sci. USA* **86**, 5044-5048.

Chalker, A. F., Leach, D. R. F. & Lloyd, R. G. (1988). *Escherichia coli* *sbcC* mutants permit the stable propagation of replicons containing a long palindrome. *Gene* **71**, 201-205.

Chédin, F., Dervyn, E., Dervyn, R., Ehrlich, S. D. & Noirot, P. (1994). Frequency of deletion formation decreases exponentially with distance between short repeats. *Mol. Microbiol.* **12**, 561-569.

Chia, W., Scott, M. R. D. & Rigby, P. W. J. (1982). The construction of cosmid libraries of eukaryotic DNA using Homer series of vectors. *Nucleic Acids Res.* **10**, 2503-2520.

Choo, K. H., Filby, G., Greco, S., Lau, Y.-F. & Kan, Y. W. (1986). Cosmid vectors for high efficiency DNA-mediated transformation and gene amplification in mamalian cells: Studies with the human growth hormone gene. *Gene* **46**, 277-286.

Collick, A., Drew, J., Penberth, J., Bois, P., Luckett, J., Scaerou, F., Jeffreys, A. & Reik, W. (1996). Instability of long inverted repeats within mouse transgenes. *EMBO* **15**, 1163-1171.

Collins, C. J., Jackson, D. A. & De Vries, F. A. J. (1976). Biochemical construction of specific chimeric plasmids from ColE1 DNA and unfractionated *Escherichia coli* DNA. *Proc. Natl. Acad. Sci. USA* **73**, 3838-3842.

Collins, J. & Hohn, B. (1978). Cosmids: a type of plasmid gene-cloning vector that is packageable *in vitro* in bacteriophage λ heads. *Proc. Natl. Acad. Sci. USA* **75**, 4242-4246.

Collins, J. (1981). Instability of palindromic DNA in *Escherichia coli*. *Cold Spring Harbor Symp. Quant. Biol.* **45**, 409-416.

Collins, J., Volkaert, G. & Nevers, P. (1982). Precise and nearly-precise excision of the symmetrical inverted repeats of Tn5; common features of *recA*-independent deletion events in *Escherichia coli*. *Gene* **19**, 139-146.

Connell, T. D., Martone, A. J. & Holmes, R. K. (1995). A new mobilizable cosmid vector for use in *Vibrio cholerae* and other gram-negative bacteria. *Gene* **153**, 85-87.

Connelly, J. C. & Leach, D. R. F. (1996). The *sbcC* and *sbcD* genes of *Escherichia coli* encode a nuclease involved in plandrome mediated inviability and genetic recombination. *Genes to Cells* **1**, 285-291.

Das Gupta, U., Weston-Hafer, K. & Berg, D. E. (1987). Local DNA sequence control of deletion formation in *Escherichia coli* plasmid pBR322. *Genetics* **115**, 41-49.

Davison, A. (1994). DNA Secondary Structure *in Vivo*. Ph.D., University of Edinburgh.

Dayn, A., Samadashwily, G. M. & Mirkin, S. M. (1992). Intramolecular DNA triplexes: unusual sequence requirements and influence on DNA polymerization. *Proc. Natl. Acad. Sci. USA* **89**, 11406-11410.

Denis, F. & Brzezinski, R. (1992). A versatile shuttle cosmid vector for use in *Escherichia coli* and *actinomycetes*. *Gene* **111**, 115-118.

De Smet, K. A. L., Jamil, S. & Stoker, N. G. (1993). Tropist3: a cosmid vector for simplified mapping of both G + C-rich and A + T-rich genomic DNA. *Gene* **136**, 215-219.

Dianov, G. L. Kuzminov, A. V. Mazin, A. V. & Salganik, R. I. (1991). Molecular mechanisms of deletion formation in *Escherichia coli* plasmids. *Mol. Gen. Genet.* **228**, 153-159

Dimri, G. P. & Das, H. K. (1990). Cloning and sequence analysis of *gyrA* gene of *Klebsiella pneumoniae*. *Nucleic Acids Res.* **15**, 151-156.

Donlon, T. A., Lalande, M., Wyman, A., Bruns, G. & Latt, S. A. (1986). Isolation of molecular probes associated with chromosome 15 instability in the Prader-Willi syndrome. *Proc. Natl. Acad. Sci. USA* **83**, 4408-4412.

Dower, W. J., Miller, J. F. & Ragsdale, C. W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* **16**, 6127.

Ehrlich, S. D., Noirot, P., Petit, M. A., Janniere, L., Michel, B. & te Riele, H. (1986). Structural instability of *Bacillus subtilis* plasmids. In *Genetic Engineering* (Setlow, J. K. & Hollander, A., eds.), Vol. 8, pp. 71-83. Plenum publishing Corp., New York.

Ehrlich, S. D. (1989). Illegitimate recombination in bacteria. In *Mobile DNA* (Berg, D. E. & Howe, M. M., eds.), pp. 799-832. Am. Soc. Microbiol., Washington DC.

Ehrlich, E., Craig, A., Poustka, A., Frischauf, A-M. & Lehrach, H. (1987). A family of cosmid vectors with the multi-copy R6K replication origin. *Gene* 57, 229-237.

Ehrlich, S. D., Bierne, H., d'Alençon, E., Villette, D., Petranovic, M., Noirot, P. & Michel, B. (1993). Mechanisms of illegitimate recombination. *Gene* 135, 161-166.

Evans, G. A. & Wahl, G. M. (1987). Cosmid vectors for genomic walking and rapid restriction mapping. In *Methods in Enzymology* (Berger, S. H. & Kimmel, A. R., eds.), Vol. 152, pp. 604-610. Academic Press Inc, Orlando.

Evans, G. A., Lewis, K. & Rothenburg, B. E. (1989). High efficiency vectors for cosmid microcloning and genomic analysis. *Gene* 79, 9-20.

Fangman, W. L & Novick, A. (1966) Mutant bacteria showing efficient utilization of thymidine. *J. Bacteriol.* 91, 2390-2391.

Franklin, N. C. (1971). Illegitimate recombination. In *The Bacteriophage Lambda* (Hershey, A. D., ed.), pp. 175-194. Cold Spring Harbour Laboratory Press, New York.

Freund, A. M., Bichara, M. & Fuchs, R. P. P. (1989). Z-DNA-forming sequences are spontaneous deletion hot spots. *Proc. Natl. Acad. Sci. USA* 86, 7465-7469.

Frey, J., Bagdasarian, M., Feiss, D., Franklin, C. H. & Deshusses, J. (1983). Stable cosmid vectors that enable the introduction of cloned fragments into a wide range of Gram-negative bacteria. *Gene* **24**, 299-308.

Gallie, D. R., Novak, S. & Kado, C. I. (1985). Novel high- and low-copy number stable cosmids for use in *Agrobacterium* and *Rhizobium*. *Plasmid* **14**, 171-175.

Gibson, T. J., Coulson, A. R., Sulston, J. E. & Little, P. F. R. (1987). Lorist2, a cosmid with transcriptional terminators insulating vector genes from interference by promoters within the insert: effect on DNA yield and cloned insert frequency. *Gene* **53**, 275-281.

Gibson, F. P., Leach, D. R. F. & Lloyd, R. G. (1992). Identification of *sbcD* mutations as cosuppressors of *recBC* that allow propagation of DNA palindromes in *Escherichia coli* K-12. *J. Bacteriol.* **174**, 1222-1228.

Gilson, E., Clement, J. M., Perrin, D. & Hofnung, M. (1990). DNA polymerase I and a protein complex bind specifically to *E. coli* palindromic unit highly repetitive DNA: Implications for bacterial chromosome organisation. *Nucleic Acids Res.* **18**, 3941-3952.

Gilson, E., Saurin, W., Perrin, D., Bachelier, S. & Hofnung, M. (1991). Palindromic units are part of a new bacterial interspersed mosaic element (BIME). *Nucleic Acids Res.* **19**, 1375-1383.

Glickman, B. W. & Ripley, L. S. (1984). Structural intermediates of deletion mutagenesis: A role for palindromic DNA. *Proc. Natl. Acad. Sci. USA* **81**, 512-516.

Gordenin, D. A., Lobachev, K. S., Degtyareva, N. P., Malkova, A. L., Perkins, E. & Resnick, M. A. (1993). Inverted DNA repeats: A source of eukaryotic genomic instability. *Mol. Cell. Biol.* **13**, 5315-5322.

Green, E. D., Riethman, H. C. Dutchik, J. E. & Olson, M. V. (1991). Detection and characterization of chimeric yeast artificial chromosome clones. *Genomics* **11**, 658-659.

Greenblatt, M. S., Grollman, A. P. & Harris, C. C. (1996). Deletions and insertions in the p53 tumor suppressor gene in human cancers: confirmation of the DNA polymerase slippage/misalignment model. *Cancer Res.* **56**, 2130-2136.

Grosveld, F. G., Lund, T., Murray, E. J., Mellor, A. L., Dahl, H. H. M. & Flavell, R. A. (1982). The constuction of cosmid libraries which can be used to transform eukaryotic cells. *Nucleic Acids Res.* **10**, 6715-6732.

Hagan, C. E. & Warren, G. J. (1983). Viability of palindromic DNA is restored by deletions at low but variable frequency in plasmids of *Escherichia coli*. *Gene* **24**, 317-326.

He, W., Wilde, A. & Kaderbhai, M. A. (1989). A simple single-step procedure for small-scale preparation of *Escherichia coli* plasmids. *Nucleic Acids Res.* **18**, 1660.

Henderson, S. T. & Petes, T. D. (1993). Instability of a plasmid-borne inverted repeat in *Saccharomyces cereviciae*. *Genetics* **113**, 57-62.

Higgins, C. F., Ammes, G. F.-L., Barnes, W. M., Clement, J. M. & Hofnung, M. (1982). A novel intercistronic regulatory element of prokaryotic operons. *Nature* **298**, 760-762.

Higgins, R. R. & Becker, A. (1995). Interaction of terminase, the DNA packaging enzyme of phage λ , with its *cos* DNA substrate. *J. Mol. Biol.* **252**, 31-46.

Hinshelwood, S. & Stoker, N. G. (1992). An *Escherichia coli*-*Mycobacterium* shuttle vector, pMSC1. *Gene* **110**, 115-118.

Hirota, Y., Yasuda, S., Yamada, M., Nisimura, A., Sugimoto, H., Sugisaki, H., Oka, A. & Takanami (1979). Structural and functional properties of the *Escherichia coli* origin of replication. *Cold Spring Harbor Symp. Quant. Biol.* **43**, 129-138.

Hite, J. M., Eckert, K. A. & Cheng, K. C. (1996). Factors affecting fidelity of DNA synthesis during PCR amplification of d(C-A)_n.d(G-T)_n microsatellite repeats. *Nucleic Acids Res.* **24**, 2429-2434.

Hohn, B. & Hohn, T. (1974). Activity of empty, headlike particles for packaging of DNA of bacteriophage λ *In Vitro. Proc. Nat. Acad. Sci. USA* **71**, 2372-2376.

Hohn, B. (1975). DNA as a substrate for packaging into bacteriophage *in vitro*. *J. Mol. Biol.* **98**, 93-106.

Hohn, B. & Collins, J. (1980). A small cosmid for efficient cloning of large DNA fragments. *Gene* **11**, 291-298.

Hohn, B. & Hinnen, A. (1980). Cloning with cosmids in *E. coli* and yeast. In *Genetic Engineering* (Setlow, J. K. & Hollaender, A., eds.), Vol. **2**, pp. 169-183. Plenum, New York.

Horwitz, J. P., Chua, J., Curby, R. J., Tomson, A. J., DaRooge, M. A., Fisher, B. E., Maurico, J. & Klundt, I. (1964). Substrates for cytochemical demonstration of enzyme activity. I. Some substituted 3-indolyl- β -D-glycopyranosides. *J. Med. Chem.* **7**, 574.

- Hulton, C. S. J., Higgins, C. F. & Sharp, P. M. (1991). ERIC sequences: A novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other *enterobacteria*. *Mol. Microbiol.* **5**, 825-834.
- Ioannou, P. A., Amemiya, C. T., Garnes, J., Kroisel, P. M., Shizuya, H., Chen, C., Batzer, M. A. & deJong, P. J. (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet.* **6**, 84-89.
- Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D. & Perucho, M. (1993). Ubiquitous somatic mutations in simple repeated sequences. *Nature* **363**, 558-561.
- Ishiura, M., Hazumi, N., Koide, T., Uchida, T. & Okada, Y. (1989). A *recB recC sbcB recJ* host prevents *recA*-independent deletions in recombinant cosmid DNA propagated in *Escherichia coli*. *J. Bacteriol.* **171**, 1068-1074.
- Ishiura, M., Ohashi, H., Uchida, T. & Okada, Y. (1989). Efficient simplified cosmid cloning: Construction and characterization of cosmid vectors that carry the two cohesive end sites of λ phages arrayed in tandem. *Anal. Biochem.* **176**, 117-127.
- Ishiura, M., Hazumi, N., Shinagawa, H., Nakata, A., Uchida, T. & Okada, Y. (1990). RecA-independent high-frequency deletion of recombinant cosmid DNA in *Escherichia coli*. *J. Gen. Microbiol.* **136**, 69-79.
- Jiang, X-M., Brahmbhatt, H. N., Quigley, N. B. & Reeves, P. R. (1987). A low copy number cosmid. *Plasmid* **18**, 170-172.
- Ji, H., Smith, L. M. & Guilfoyle, R. A. (1994). Rapid isolation of cosmid insert DNA by triple-helix-mediated affinity capture. *Genet. Anal. Tech. Appl.* **11**, 43-47.

Ji, H., Francisco, T., Smith, L. M. & Guilfoyle, R. A. (1996). Rapid restriction mapping of cosmids by sequence-specific triple-helix-mediated affinity capture. *Genomics* **31**, 185-192.

Kim, U.-J., Shizuya, H., de Jong, P. J., Birren, b. & Simon, M. (1992). Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Res.* **20**, 1083-1085.

Klein, H. L. (1995). Genetic control of intrachromosomal recombination. *BioEssays* **17**, 147-159.

Klysik, J., Stirdivant, S. M. & Wells, R. D. (1982). Left handed DNA: cloning, characterization, and instability of inserts containing different lengths of (dC-dG) in *Escherichia coli*. *J. Biol. Chem.* **257**, 10152-10158.

Kowalczykowski, S. C., Dixon, D. D., Eggleston, A. K., Lauder, S. D. & Rehrauer, W. M. (1994). Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol. Rev.* **58**, 401-465.

Kurnit, D. M. (1989). *Escherichia coli recA* deletion strains that are highly competent for transformation and for *in vivo* phage packaging. *Gene* **82**, 313-315.

Kushner, S. R. (1978). An improved method for transformation of *E. coli* with ColE1 derived plasmids. In *Genetic Engineering* (Boyer, H. W. & Nicosia, S., eds.), pp. 17-23. Elsevier/North-Holland Biomedical Press, Amsterdam.

Kuzminov, A. V. (1996). Mutant fixation via plasmid dimerization and its relation to human diseases. *TIG* **12**, 246-249.

L'Abbe, D., Duhaime, J., Lang, B. F. & Morais, R. (1991). The transcription of DNA in chicken mitochondria initiates from one major promoter. *J. Biol. Chem.* **266**, 10844-10850.

LaDuca, R. J., Fay, P. J., Chuang, C., McHenry, C. S. & Bambara, R. A. (1983). Site-specific pausing of deoxyribonucleic acid synthesis catalyzed by four forms of *Escherichia coli* DNA polymerase III. *Biochemistry* **22**, 5177-5188.

Leach, D. R. F. & Stahl, F. W. (1983). Viability of lambda phages carrying a perfect palindrome in the absence of recombination nucleases. *Nature* **305**, 448-451.

Leach, D. R. F. (1994). Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. *BioEssays* **16**, 893-900.

Leach, D. R. F. (1996). *Genetic Recombination*, Blackwell Science, Oxford.

Lin, L. & Meyer, R. J. (1987). DNA synthesis is initiated at two positions within the origin of replication of plasmid R1162. *Nucleic Acids Res.* **15**, 8319-8331.

Lindsey, J. C. & Leach, D. R. F. (1988). Slow replication of palindrome containing DNA. *J. Mol. Biol.* **206**, 779-782.

Little, P. R. F. & Cross, S. H. (1985). A cosmid vector that facilitates restriction enzyme mapping. *Proc. Natl. Acad. Sci. USA* **82**, 3159-3163.

Little, P. F. R. (1989). Choice and use of cosmid vectors. In *DNA cloning volume III: a practical approach* (Glover, D. M., ed.), Vol. **3**, pp. 19-42. IRL Press, Oxford.

- Lovett, S. T., Drapkin, P. T., Sutter, V. A. J. & Gluckman-Peskind, T. J. (1993). A sister-strand exchange mechanism for recA-independent deletion of repeated DNA sequences in *Escherichia coli*. *Genetics* **135**, 631-642.
- Mandel, M. & Higa, A. (1970). Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* **53**, 159-162.
- Matsuoka, T., Kato, H., Hashimoto, K. & Kurosawa, Y. (1991). Direct cloning of a long restriction fragment aided with a jumping clone. *Gene* **107**, 27-35.
- Matfield, M., Badawi, R. & Brammer, W. J. (1985). Rec-dependent and Rec-independent recombination of plasmid-borne duplications in *Escherichia coli*. *Mol. Gen. Genet.* **199**, 518-523.
- Mazin, A. V., Kuzminov, A. V., Dianov, G. L. & Salganik, R. I. (1991). Mechanisms of deletion formation in *Escherichia coli* plasmids. *Mol. Gen. Genet.* **228**, 209-214.
- Meijer, M., Beck, E., Hansen, F. G., Bergmans, H. E. N., Messer, W., von Meyenburg, K. & Scaller (1979). Nucleotide sequence of the origin of replication of the *Escherichia coli* K-12 chromosome. *Proc. Natl. Acad. Sci. USA* **76**, 580-584.
- Meuth, M. (1989). Illegitimate recombination in mammalian cells. In *Mobile DNA* (Berg, D. E. & Howe, M. M., eds.), pp. 833-860. American Society for Microbiology, Washington DC.
- Meuth, M. (1990). The structure of mutation in mammalian cells. *Biochim. Biophys. Acta* **1032**, 1-17.
- Mizuuchi, K., Mizuuchi, M. & Gellert, M. (1982). Cruciform structures in palindromic DNA are favoured by DNA supercoiling. *J. Mol. Biol.* **156**, 229-243.

Müller, U. R. & Fitch, W. M. (1982). Evolutionary selection for perfect hairpin structures in viral DNAs *Nature* **298**, 582-585.

Murialdo, H. (1991). Bacteriophage Lambda DNA maturation and Packaging. *Annu. Rev. Biochem.* **60**, 125-153.

Murchie, A. I. J. & Lilley, D. M. J. (1992). Supercoiled DNA and cruciform structures. In *DNA structures partA: Synthesis and physical analysis of DNA* (Lilley, D. M. J. & Dahlberg, J. E., eds.), Vol. **211**, pp.158-180. Methods in Enzymology, Academic Press Inc., San Diego.

Nakano, Y., Yoshida, Y., Yamahita, Y. & Koga, T. (1995). Construction of a series of pACYC-derived plasmid vectors. *Gene* **162**, 157-158.

Naom, I. S., Morton, S. J., Leach, D. R. F. & Lloyd, R. G. (1989). Molecular organisation of *sbcC*, a gene that affects genetic recombination and the viability of DNA palindromes in *Escherichia coli* K-12. *Nucleic Acids Res.* **17**, 8033-8045.

Neil, D. L., Villasante, A., Fisher, R. B., Vetrie, D., Cox, B. & Tyler-Smith, C. (1990). Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. *Nucleic Acids Res.* **18**, 1421-1428.

Ohno, S. (1970). *Evolution by gene duplication*, Springer-Verlag, New york.

Oppenheim, A. B., Rudd, K. E., Mendelson, I. & Teff, D. (1993) integration host factor binds to a unique class of complex repetitive extragenic DNA sequences in *Escherichia coli*. *Mol. Microbiol.* **10**, 113-122.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. & Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as Single-Stranded Conformation Polymorphisms. *Proc. Natl. Acad. Sci. USA* **86**, 2766-2770.

Palva, E. T., Liljeström, P. & Harayama, S. (1981). Cosmid cloning and transposon mutagenesis in *Salmonella typhimurium* using phage λ vehicles. *Mol. Gen. Genet.* **181**, 153-157.

Park, C., Campbell, J. L. & Goddard, W. A. (1993). Design superiority of palindromic DNA sites for site-specific recognition of proteins: Tests using protein stitchery. *Proc. Natl. Acad. Sci. USA* **90**, 4892-4896.

Peeters, B. P. H., deBoer, J. H., Bron, S. & Venema, G. (1988). Structural plasmid instability in *Bacillus subtilis*: Effect of direct and inverted repeats. *Mol. Gen. Genet.* **212**, 450-458.

Pirrotta, V., Hadfield, C. & Pretorius, G. H. J. (1983). Microdissection and cloning of the *white* locus and the 3B1-3C2 region of the *Drosophila* X chromosome. *EMBO J.* **2**, 927-934.

Platt, J. R. (1986). Transcriptional termination and the regulation of gene expression. *Annu. Rev. Biochem.* **55**, 339-372.

Poulton, J., Deadman, M. E., Bindoff, L., Morten, K., Land, J. & Brown, G. (1993). Families of mtDNA re-arrangements can be detected in patients with mtDNA deletions: duplications may be a transient intermediate form. *Hum. Mol. Genet.* **2**, 23-30.

Primrose, S. B. & Ehrlich, S. D. (1981). Isolation of plasmid deletion mutants and study of their instability. *Plasmid* **6**, 193-201.

Rackwitz, H. R., Zehetner, G. Murialdo, H., Delius, H., Chia, J. H., Poustka, A., Frischauf, A-M. & Lehrach, H. (1985). Analysis of cosmids using linearization by phage lambda terminase. *Gene* **40**, 259-266.

Rao, B. S., Manor, H. & Martin, R. G. (1988). Pausing in simian virus 40 DNA replication by a sequence containing (dG-dA)₂₇.(dT-dC)₂₇. *Nucleic Acids Res.* **16**, 8077-8094.

Richards, R. I. & Sutherland, G. R. (1992). Dynamic mutations: A new class of mutations causing human disease. *Cell* **70**, 709-712.

Ruskin, B. & Fink, G. R. (1993). Mutations in POL1 increase the mitotic instability of tandem inverted repeats in *Saccharomyces cerevisiae*. *Genetics* **133**, 43-56.

Sadowski. (1993). Site-specific genetic recombination: hops, flips, and flops. *FASEB J.* **7**, 760-767.

Saito, I. & Stark, G. R. (1986). Charomids: Cosmid vectors for efficient cloning and mapping of large or small restriction fragments. *Proc. Natl. Acad. Sci. USA* **83**, 8664-8668.

Samadashwily, G. M., Dayn, A. & Mirkin, S. M. (1993). Suicidal nucleotide sequences for DNA polymerization. *EMBO J.* **12**, 4975-4983.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.

Schlessinger, D. (1990). Yeast artifical chromosomes: tools for mapping and analysis of complex genomes. *TIG* **6**, 248-258.

Shafferman, A., Flashner, Y., Hertman, I., Olami, Y. & Cohen, S. (1987). Molecular aspects of genomic instability of an artificial 68 bp perfect palindrome in *Escherichia coli*. *Mol. Gen. Genet.* **208**, 294-300.

Sharples, G. J. & Lloyd, R. G. (1990). A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. *Nucleic Acids Res.* **18**, 6053-6058.

Shimizu, M., Gellibolian, R., Oostra, B. A. & Wells, R. D. (1996). Cloning, characterization and properties of plasmids containing CGG triplet repeats from the FMR-1 gene. *J. Mol. Biol.* **258**, 614-626.

Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri Y. & Simon M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector.

Proc. Natl. Acad. Sci. USA **89**, 8794-8797.

Shurvinton, C. E., Stahl, M. M. & Stahl, F. W. (1987). Large palindromes in the Lambda phage genome are perserved in a rec^+ host by inhibiting Lambda DNA replication. *Proc. Natl. Acad. Sci. USA* **84**, 1624-1628.

Sinden, R. R. & Wells, R. D. (1992). DNA structure, mutations and human disease. *Cur. Opin. Biotechnol.* **3**, 612-622.

Smith, G. R. (1987). Homologous recombination in prokaryotes. *Microbiol. Rev.* **52**, 1-28.

Soeda, E. (1993). Structural analysis of human genome by YAC technologies. *Nippon Rinsho* **51**, 2246-2251.

Speek, M., Raff, J. W., Harrison-Lavoie, K., Little, P. R. F. & Glover, D. M. (1988). Smart2, a cosmid vector with a phage lambda origin for both systematic chromosome walking and P-element-mediated gene transfer in *Drosophila*. *Gene* **64**, 173-177.

Spoerel, N. A. & Kafatos, F. C. (1987). Isolation of full-length genes: Walking the chromosome. In *Methods in Enzymology* (Berger, S. H. & Kimmel, A. R., eds.), Vol. **152**, pp. 598-603. Academic Press Inc, Orlando.

Stallings, R. L., Torney, D. C., Hildebrand, C. E., Longmire, J. L., Deaven, L. L., Jett, J. H., Doggett, N. A. & Moyzis, R. K. (1990). Physical mapping of human chromosomes by repetitive sequence fingerprinting. *Proc. Natl. Acad. Sci. USA* **87**, 6218-6222.

Stark, G. R., Debatisse, M., Giulotto, E. & Wahl, G. M. (1989). Recent progress in the understanding of mechanisms of mammalian DNA amplification. *Cell* **57**, 901-908.

Steinmetz, M., Winoto, A., Minard, K. & Hood, L. (1982). Clusters of genes encoding mouse transplantation antigens. *Cell* **28**, 489-498.

Streisinger, G., Okada, Y., Emerich, J., Newton, J., Tsugita, A., Terzaghi, E. & Inouye, M. (1960). Frame shift mutations and genetic code. *Cold Spring Harbor Symp. Quant. Biol.* **31**, 77-84.

Summers, D. K. (1994). The origins and consequences of genetic instability in prokaryotes. *Dev. Biol. Stand.* **83**, 7-11.

Thibodeau, S. N., Bren, G. & Schaid, D. (1993). Microsatellite instability in cancer of the proximal colon. *Science* **260**, 816-819.

Trinh, T. Q. & Sinden, R. R. (1993). The influence of primary and secondary DNA structure in deletion and duplication between direct repeats in *Escherichia coli*. *Genetics* **134**, 409-422.

Usdin, K. & Woodford, K. J. (1995). CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis *in vitro*. *Nucleic Acids Res.* **23**, 4202-4209.

Vieira, J. & Messing, J. (1982). The pUC plasmids, an M13mp7 derived system for insertional mutagenesis and sequencing with universal primers. *Gene* **19**, 259-268.

Vieira, J. & Messing, J. (1991). New pUC-derived cloning vectors with different selectable markers and DNA replication origins. *Gene* **100**, 189-194.

Vincze, E. & Kiss, G. B. (1990). A phosphate group at the *cos* ends of phage lambda DNA is not a prerequisite for *in vitro* packaging: an alternative method for constructing genomic libraries using a new phasmid vector, λ pGY97. *Gene* **96**, 17-22.

Vogelstein, B. & Gillespie, D. (1979). Preparative and analytical purification of DNA from agarose. *Proc. Natl. Acad. Sci. USA* **76**, 615-619.

Wahl, G. M., Lewis, K. A., Ruiz, J. C., Rothenberg, B., Zhao, J. & Evans, G. A. (1987). Cosmid vectors for rapid genomic walking, restriction mapping, and gene transfer. *Proc. Natl. Acad. Sci. USA* **84**, 2160-2164.

Wallace, D. C., Lott, M. T., Torroni, A. & Shoffner, J. M. (1991). Report of the Committee on Human Mitochondrial DNA. *Cytogenet. Cell. Genet.* **58**, 1103-1123.

Wallace, D. C. (1992). Mitochondrial genetics: a paradigm for aging and degenerative diseases? *Science* **256**, 628-632.

Walz, A. & Pirrotta, V. (1975). Sequence of the Pr promoter of phage lambda. *Nature* **254**, 118-121.

Wang, P. Projan, S. J., Henriquez, V. & Novick, R. P. (1993). Origin recognition specificity in pT181 plasmids is determined by a functionally asymmetric palindrome DNA element. *EMBO J.* **12**, 45-52.

Warren, G. J. & Green, R. L. (1985). Comparison of physical and genetic properties of palindromic DNA sequences. *J. Bacteriol.* **161**, 1103-1111.

Weaver, D. T. & DePamphilis, M. L. (1984). The role of palindromic and non-palindromic sequences in arresting DNA synthesis *in vitro* and *in vivo*. *J. Mol. Biol.* **180**, 961-986.

Wells, R. D. & Sinden, R. R. (1993). Defined ordered sequence DNA, DNA structure, and DNA directed mutation. In *Genome analysis* (Davies, D. & Warren, S., eds.), Vol. 7, pp. 107-138. Cold Spring Harbour laboratory Press, New York.

Wells, R. D. (1996). Molecular basis of genetic instability of triplet repeats. *J. Biol. Chem.* **271**, 2875-2878.

West, S. C. (1992). Enzymes and molecular mechanisms of genetic recombination. *Annu. Rev. Biochem.* **61**, 603-640.

Weston-Hafer, C. & Berg, D. E. (1991). Deletions in plasmid pBR322: Replication slippage involving leading and lagging strands. *Genetics* **127**, 649-655.

Wilson, D. A. & Thomas, C. A. (1974). Palindromes in chromosomes. *J. Mol. Biol.* **84**, 115-144.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.

Yang, Y. & Ames, G. F.-L. (1990). The family of repetitive extragenic palindromic sequences: Interaction with DNA gyrase and histonelike protein HU. In *The Bacterial Chromosome* (Drlica, K. & Ripley, M., eds), pp.211-225. American Society for Microbiology, Washington DC.

Yelton, M. M., Timberlake, W. E. & Van Den Hondel, C. A. M. J. J. (1985). A cosmid for selecting genes by complementation in *Aspergillus nidulans*: selection of the developmentally regulated *yA* locus. *Proc. Natl. Acad. Sci. USA* **82**, 834-838.

Yoshimura, H., Yoshino, T., Hirose, T., Nakamura, Y., Higashi, M., Hase, T., Yamaguchi, K., Hirokawa, H. & Masamune, Y. (1986). Biological characteristics of palindromic DNA (II). *J. Gen. Appl. Microbiol.* **32**, 393-404.

Zhou, X., Deng, Z., Hopwood, D. A. & Kieser, T. (1994). Characterization of ϕ HAU3, a broad-host-range temperature *Streptomyces* phage, and development of phasmids. *J. Bacteriol.* **176**, 2096-2099.

APPENDIX 1

Sequence of cDrII; 5` — 3` DNA; Length: 4885bp.

```

1  GCGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT

51  GCAGCTGGCA CGACAGGTTT CCCGACTGGA AAGCGGGCAG TGAGCGCAAC

101 GCAATTAATG TGAGTTAGCT CACTCATTAG GCACCCAGG CTTTACACTT

151 TATGCTTCCG GCTCGTATGT TGTGTGGAAT TGTGAGCGGA TAACAATTTC

201 ACACAGGAAA CAGCTATGAC CATGATTACG AATTCGAGCT CGGTACCCGG
      GAATTC (EcoRI)
251 GGATCCTCTA GAGTCGACCT GCAGGCATGC AAGCTTGGCA CTGGCCGTGC
Multicloning Site (pUC18)      AAGCTT (HindIII)
301 TTTTACAACG TCGTGACTGG GAAAACCCTG GCGTTACCCA ACTTAATCGC

351 CTTGCAGCAC ATCCCCCTTT CGCCAGCTGG CGTAATAGCG AAGAGGCCCG

401 CACCGATCGC CCTTCCCAAC AGTTGCGCAG CCTGAATGGC GAATGGCGCC

451 TGATGCGGTA TTTTCTCCTT ACGCATCTGT GCGGTATTTC ACACCGCATA

501 TGGTGCACTC TCAGTACAAT CTGCTCTGAT GCCGCATAGT TAAGCCAGCC

551 CCGACACCCG CCAACACCCG CTGACGCGCC CTGACGGGCT TGTCTGCTCC

601 CGGCATCCGC TTACAGACAA GCTGTGACCG TCTCCGGGAG CTGCATGTGT

651 CAGAGGTTTT CACCGTCATC ACCGAAACGC GCGAGACGAA AGGCGCTCGT

701 GATACGCCTA TTTTATAGG TTAATGTCAT GATAATAATG GTTCTTAGA
      +
751 CGTTCACTTT ACGGGTCCTT TCCGGTGATC CGACGGTTAC GGGGCGGCGA
      ++++++OLIGO 1+++++-----
801 CCTCGCGGGT TTTCGCTATT TATGAAAATT TTCCGGTTTA AGGCGTTTCC
      -----
851 GTTCTTCTTC GTCATAACTT AATGTTTTTA TTTAAAATAC CCTCTGAAAA
      -----Lambda 431bp Fragment-----
901 GAAAGGAAAC GACAGGTGCT GAAAGCGAGG CTTTTTGCC TCTGTCGTTT
      -----

```



```

951 CCTTCTCTG TTTTGTCCG TGAATGAAC AATGGAAGTC AACAAAAAGC
-----
1001 AGCTGGCTGA CATTTTCGGT GCGAGTATCC GTACCATTCA GAACTGGCAG
-----
1051 GAACAGGGAA TGCCCGTTCT GCGAGGCGGT GGCAAGGGTA ATGAGGTGCT
-----
1101 TTATGACTCT GCCGCCGTCA TAAAATGGTA TGCCGAAAGG GATGCTGAAA
-----
1151 TTGAGAACGA AAAGCTGCGC CGGGAGGTTG AAGAACTGCG GCAGGCCAGC
-----
1201 GAGGCAGATC AGATCCGTCG ACCTGCAGGG GGGGGGGGGC GCTGAGGTCT
-----*****
1251 GCCTCGTGAA GAAGGTGTTG CTGACTCATA CCAGGCCTGA ATCGCCCCAT
*****
1301 CATCCAGCCA GAAAGTGAGG GAGCCACGGT TGATGAGAGC TTTGTTGTAG
*****Kanamycin Genblock*****
1351 GTGGACCAGT TGGTGATTTT GAACTTTTGC TTTGCCACGG AACGGTCTGC
*****
1401 GTTGTCGGGA AGATGCGTGA TCTGATCCTT CAACTCAGCA AAAGTTCGAT
*****
1451 TTATTCAACA AAGCCGCCGT CCCGTCAAGT CAGCGTAATG CTCTGCCAGT
*****
1501 GTTACAACCA ATTAACCAAT TCTGATTAGA AAAACTCATC GAGCATCAAA
*****
1551 TGAAACTGCA ATTTATTCAT ATCAGGATTA TCAATACCAT ATTTTGTGAA
*****
1601 AAGCCGTTTC TGTAATGAAG GAGAAACTC ACCGAGGCAG TTCATAGGA
*****
1651 TGGCAAGATC CTGGTATCGG TCTGCGATTG CACTCGTCC AACATCAATA
*****
1701 CAACCTATTA ATTTCCCCTC GTCAAAAATA AGGTATCAA GTGAGAAATC
*****
1751 ACCATGAGTG ACGACTGAAT CCGGTGAGAA TGGCAAAAGC TTATGCATTT
*****
1801 CTTTCCAGAC TTGTTCAACA GGCCAGCCAT TACGCTCGTC ATCAAAATCA
*****
1851 CTCGCATCAA CCAAACCGTT ATTCATTCTG GATTGCGCCT GAGCGAGACG
*****
1901 AAATACGCGA TCGCTGTTAA AAGGACAATT ACAAACAGGA ATCGAATGCA
*****
1951 ACCGGCGCAG GAACACTGCC AGCGCATCAA CAATATTTTC ACCTGAATCA
*****

```


2001 GGATATTCTT CTAATACCTG GAATGCTGTT TTCCCGGGGA TCGCAGTGGT

 2051 GAGTAACCAT GCATCATCAG GAGTACGGAT AAAATGCTTG ATGGTCGGAA

 2101 GAGGCATAAA TTCCGTCAGC CAGTTTAGTC TGACCATCTC ATCTGTAACA

 2151 TCATTGGCAA CGCTACCTTT GCCATGTTTC AGAAACAACCT CTGGCGCATC

 2201 GGGCTTCCCA TACAATCGAT AGATTGTCGC ACCTGATTGC CCGACATTAT

 2251 CGCGAGCCCA TTTATACCCA TATAAATCAG CATCCATGTT GGAATTTAAT

 2301 CGCGGCCTCG AGCAAGACGT TTCCCGTTGA ATATGGCTCA TAACACCCCT

 2351 TGTATTACTG TTTATGTAAG CAGACAGTTT TATTGTTTAT GATGATATAT

 2401 TTTTATCTTG TGCAATGTAA CATCAGAGAT TTTGAGACAC AACGTGGCTT

 2451 TCCCCCCCCC CCCTGCAGGT CGACGGATCT CTCGATCACT TTACGGGTCC
 *****+#####Oligo 2####
 2501 TTTCCGTTGA TCCGACGGTT ACGGGGCGGC GACCTCGCGG GTTTTCGCTA
 #####-
 2551 TTTATGAAA TTTTCCGGTT TAAGGCGTTT CCGTTCTTCT TCGTCATAAC

 2601 TTAATGTTTT TATTTAAAT ACCCTCTGAA AAGAAAGGAA ACGACAGGTG
 -----Lambda 431bp Fragment-----
 2651 CTGAAAGCGA GGCTTTTTTG CCTCTGTCGT TTCCTTTCTC TGTTTTTGTC

 2701 CGTGGAATGA ACAATGGAAG TCAACAAAAA GCAGCTGGCT GACATTTTCG

 2751 GTGCGAGTAT CCGTACCATT CAGAACTGGC AGGAACAGGG AATGCCCGTT

 2801 CTGCGAGGCG GTGGCAAGGG TAATGAGGTG CTTTATGACT CTGCCGCCGT

 2851 CATAAAATGG TATGCCGAAA GGGATGCTGA AATTGAGAAC GAAAAGCTGC

 2901 GCCGGGAGGT TGAAGAACTG CGGCAGGCCA GCGAGGCAGA TCATCGAGTA
 -----#####
 2951 CGTAGGTGGC ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT
 +++
 3001 TTTTCTAAAT ACATTCAAAT ATGTATCCGC TCATGAGACA ATAACCCTGA

3051 TAAATGCTTC AATAATATTG AAAAAGGAAG AGTATGAGTA TTCAACATTT
 3101 CCGTGTCGCC CTTATTCCCT TTTTTCGGC ATTTTGCCTT CCTGTTTTTG
 3151 CTCACCCAGA AACGCTGGTG AAAGTAAAAG ATGCTGAAGA TCAGTTGGGT
 3201 GCACGAGTGG GTTACATCGA ACTGGATCTC AACAGCGGTA AGATCCTTGA
 3251 GAGTTTTTCG CCCGAAGAAC GTTTTCCAAT GATGAGCACT TTTAAAGTTC
 3301 TGCTATGTGG CGCGGTATTA TCCCGTATTG ACGCCGGGCA AGAGCAACTC
 3351 GGTCGCCGCA TACACTATTTC TCAGAATGAC TTGGTTGAGT ACTCACCAGT
 3401 CACAGAAAAG CATCTTACGG ATGGCATGAC AGTAAGAGAA TTATGCAGTG
 3451 CTGCCATAAC CATGAGTGAT AACACTGCGG CCAACTTACT TCTGACAACG
 3501 ATCGGAGGAC CGAAGGAGCT AACCGCTTTT TTGCACAACA TGGGGGATCA
 3551 TGTAACTCGC CTTGATCGTT GGAACCGGA GCTGAATGAA GCCATACCAA
 3601 ACGACGAGCG TGACACCACG ATGCCTGTAG CAATGGCAAC AACGTTGCGC
 3651 AAACTATTAA CTGGCGAACT ACTTACTCTA GCTTCCCGGC AACAATTAAT
 3701 AGACTGGATG GAGGCGGATA AAGTTGCAGG ACCACTTCTG CGCTCGGCCC
 3751 TTCCGGCTGG CTGGTTTATT GCTGATAAAT CTGGAGCCGG TGAGCGTGGG
 3801 TCTCGCGGTA TCATTGCAGC ACTGGGGCCA GATGGTAAGC CCTCCCGTAT
 3851 CGTAGTTATC TACACGACGG GGAGTCAGGC AACTATGGAT GAACGAAATA
 3901 GACAGATCGC TGAGATAGGT GCCTCACTGA TTAAGCATTG GTAAGTGTCA
 3951 GACCAAGTTT ACTCATATAT ACTTTAGATT GATTAAAAC TTCATTTTAA
 4001 ATTTAAAAGG ATCTAGGTGA AGATCCTTTT TGATAATCTC ATGACCAAAA
 4051 TCCCTTAACG TGAGTTTTTCG TTCCACTGAG CGTCAGACCC CGTAGAAAAG

4101 ATCAAAGGAT CTTCTTGAGA TCCTTTTTTT CTGCGCGTAA TCTGCTGCTT
 4151 GCAAACAAAA AAACCACCGC TACCAGCGGT GGTTTGTGTTG CCGGATCAAG
 4201 AGCTACCAAC TCTTTTTCCG AAGGTAAGTG GCTTCAGCAG AGCGCAGATA
 4251 CCAAATACTG TCCTTCTAGT GTAGCCGTAG TTAGGCCACC ACTTCAAGAA
 4301 CTCTGTAGCA CCGCCTACAT ACCTCGCTCT GCTAATCCTG TTACCAGTGG
 4351 CTGCTGCCAG TGGCGATAAG TCGTGCTTA CCGGGTTGGA CTCAAGACGA
 4401 TAGTTACCGG ATAAGGCGCA GCGGTCGGGC TGAACGGGGG GTTCGTGCAC
 4451 ACAGCCCAGC TTGGAGCGAA CGACCTACAC CGAACTGAGA TACCTACAGC
 4501 GTGAGCTATG AGAAAGCGCC ACGCTTCCCG AAGGGAGAAA GGCGGACAGG
 4551 TATCCGGTAA GCGGCAGGGT CGGAACAGGA GAGCGCACGA GGGAGCTTCC
 4601 AGGGGGAAAC GCCTGGTATC TTTATAGTCC TGTCGGGTTT CGCCACCTCT
 4651 GACTTGAGCG TCGATTTTGG TGATGCTCGT CAGGGGGGCG GAGCCTATGG
 4701 AAAAACGCCA GCAACGCGGC CTTTTTACGG TTCCTGGCCT TTTGCTGGCC
 4751 TTTTGCTCAC ATGTTCTTTC CTGCGTTATC CCCTGATTCT GTGGATAACC
 4801 GTATTACCGC CTTTGAGTGA GCTGATACCG CTCGCCGAG CCGAACGACC
 4851 GAGCGCAGCG AGTCAGTGAG CGAGGAAGCG GAAGA

Sequence of cDrIII; 5' — 3' DNA; Length: 4885bp.

1 GCGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT
 51 GCAGCTGGCA CGACAGGTTT CCCGACTGGA AAGCGGGCAG TGAGCGCAAC
 101 GCAATTAATG TGAGTTAGCT CACTCATTAG GCACCCAGG CTTTACACTT


```

151 TATGCTTCCG GCTCGTATGT TGTGTGGAAT TGTGAGCGGA TAACAATTC

201 ACACAGGAAA CAGCTATGAC CATGATTACG AATTCGAGCT CGGTACCCGG
      GAATTC (EcoRI)

251 GGATCCTCTA GAGTCGACCT GCAGGCATGC AAGCTTGGCA CTGGCCGTGC
Multicloning Site (pUC18)      AAGCTT (HindIII)

301 TTTTACAACG TCGTGA CTGG GAAAACCCCTG GCGTTACCCA ACTTAATCGC

351 CTTGCAGCAC ATCCCCCTTT CGCCAGCTGG CGTAATAGCG AAGAGGCCCG

401 CACCGATCGC CCTTCCCAAC AGTTGCGCAG CCTGAATGGC GAATGGCGCC

451 TGATGCGGTA TTTTCTCCTT ACGCATCTGT GCGGTATTTC ACACCGCATA

501 TGGTGCACTC TCAGTACAAT CTGCTCTGAT GCCGCATAGT TAAGCCAGCC

551 CCGACACCCG CCAACACCCG CTGACGCGCC CTGACGGGCT TGTCTGCTCC

601 CGGCATCCGC TTACAGACAA GCTGTGACCG TCTCCGGGAG CTGCATGTGT

651 CAGAGGTTTT CACCGTCATC ACCGAAACGC GCGAGACGAA AGGCGCTCGT

701 GATACGCCTA TTTTATAGG TTAATGTCAT GATAATAATG GTTCTTAGA
      +
751 CGTTCAC TTT ACGGGTCCTT TCCGGTGATC CGACGGTTAC GGGGCGGCGA
      ++++++OLIGO 1+++++-----
801 CCTCGCGGGT TTTCGCTATT TATGAAAATT TTCCGGTTTA AGGCGTTTCC
      -----
851 GTTCTTCTTC GTCATAACTT AATGTTTTTA TTAAAATAC CCTCTGAAAA
      -----Lambda 431bp Fragment-----
901 GAAAGGAAAC GACAGGTGCT GAAAGCGAGG CTTTTTGGCC TCTGTCGTTT
      -----
951 CCTTTCTCTG TTTTGTCCG TGAATGAAC AATGGAAGTC AACAAAAAGC
      -----
1001 AGCTGGCTGA CATTTTCGGT GCGAGTATCC GTACCATTCA GAACTGGCAG
      -----
1051 GAACAGGGAA TGCCCGTTCT GCGAGGCGGT GGCAAGGGTA ATGAGGTGCT
      -----
1101 TTATGACTCT GCCGCCGTCA TAAAATGGTA TGCCGAAAGG GATGCTGAAA
      -----
1151 TTGAGAACGA AAAGCTGCGC CGGGAGGTTG AAGAACTGCG GCAGGCCAGC
      -----

```



```

1201 GAGGCAGATC AGATCCGTCG ACCTGCAGGG GGGGGGGGGA AAGCCACGTT
-----*****
1251 GTGTCTCAAA ATCTCTGATG TTACATTGCA CAAGATAAAA ATATATCATC
*****
1301 ATGAACAATA AACTGTCTG CTTACATAAA CAGTAATACA AGGGGTGTTA
*****Kanamycin Genblock*****
1351 TGAGCCATAT TCAACGGGAA ACGTCTTGCT CGAGGCCGCG ATTAAATTCC
*****
1401 AACATGGATG CTGATTTATA TGGGTATAAA TGGGCTCGCG ATAATGTCGG
*****
1451 GCAATCAGGT GCGACAATCT ATCGATTGTA TGGGAAGCCC GATGCGCCAG
*****
1501 AGTTGTTTCT GAAACATGGC AAAGGTAGCG TTGCCAATGA TGTACAGAT
*****
1551 GAGATGGTCA GACTAACTG GCTGACGGAA TTTATGCCTC TTCCGACCAT
*****
1601 CAAGCATTTT ATCCGTACTC CTGATGATGC ATGGTTACTC ACCACTGCGA
*****
1651 TCCCCGGGAA AACAGCATT CAGGTATTAG AAGAATATCC TGATTCAGGT
*****
1701 GAAAATATTG TTGATGCGCT GGCAGTGTTT CTGCGCCGGT TGCATTGAT
*****
1751 TCCTGTTTGT AATGTCTCTT TTAACAGCGA TCGCGTATTT CGTCTCGCTC
*****
1801 AGGCGCAATC ACGAATGAAT AACGGTTTGG TTGATGCGAG TGATTTTGAT
*****
1851 GACGAGCGTA ATGGCTGGCC TGTGTAACAA GTCTGGAAAG AAATGCATAA
*****
1901 GCTTTTGCCA TTCTACCGG ATTCAGTCGT CACTCATGGT GATTCTCAC
*****
1951 TTGATAACCT TATTTTGTAC GAGGGGAAAT TAATAGGTTG TATTGATGTT
*****
2001 GGACGAGTCG GAATCGCAGA CCGATACCAG GATCTTGCCA TCCTATGGAA
*****
2051 CTGCCTCGGT GAGTTTCTC CTTCATTACA GAAACGGCTT TTTCAAAAAT
*****
2101 ATGGTATTGA TAATCTGAT ATGAATAAAT TGCAGTTTCA TTTGATGCTC
*****
2151 GATGAGTTT TCTAATCAGA ATTGGTTAAT TGGTTGTAAC ACTGGCAGAG
*****

```


2201 CATTACGCTG ACTTGACGGG ACGGCGGCTT TGTGAATAA ATCGAACTTT

 2251 TGCTGAGTTG AAGGATCAGA TCACGCATCT TCCCGACAAC GCAGACCGTT

 2301 CCGTGGCAAA GCAAAAGTTC AAAATCACCA ACTGGTCCAC CTACAACAAA

 2351 GCTCTCATCA ACCGTGGCTC CCTCACTTC TGGCTGGATG ATGGGGCGAT

 2401 TCAGGCCTGG TATGAGTCAG CAACACCTTC TTCACGAGGC AGACCTCAGC

 2451 GCCCCCCCC CCCTGCAGGT CGACGATCT CTCGATCACT TTACGGGTCC
 *****+++++#####Oligo 2####
 2501 TTTCCGGTGA TCCGACGGT ACGGGGCGG GACCTCGCGG GTTTTCGCTA
 #####-----
 2551 TTTATGAAAA TTTTCCGGT TAAGGCGTTT CCGTCTTCT TCGTCATAAC

 2601 TTAATGTTTT TATTTAAAT ACCCTCTGAA AAGAAAGGAA ACGACAGGTG
 ----- Lambda 431bp Fragment-----
 2651 CTGAAAGCGA GGCTTTTGG CCTCTGTCGT TTCCTTCTC TGTTTTGTG

 2701 CGTGAATGA ACAATGGAAG TCAACAAAA GCAGCTGGCT GACATTTTCG

 2751 GTGCGAGTAT CCGTACCATT CAGAACTGGC AGGAACAGG AATGCCCGTT

 2801 CTGCGAGGCG GTGGCAAGG TAATGAGGTG CTTTATGACT CTGCCGCCGT

 2851 CATAAATGG TATGCCGAAA GGGATGCTGA AATTGAGAAC GAAAAGCTGC

 2901 GCCGGGAGGT TGAAGAACTG CGGCAGGCCA GCGAGGCAGA TCATCGAGTA
 -----#####+++
 2951 CGTAGGTGGC ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT
 +++
 3001 TTTTCTAAAT ACATTCAAAT ATGTATCCGC TCATGAGACA ATAACCCTGA

 3051 TAAATGCTTC AATAATATTG AAAAAGGAAG AGTATGAGTA TTCAACATTT

 3101 CCGTGTCGCC CTTATTCCCT TTTTTCGGC ATTTTCCTT CCTGTTTTG

 3151 CTCACCCAGA AACGCTGGTG AAAGTAAAAG ATGCTGAAGA TCAGTTGGGT

 3201 GCACGAGTGG GTTACATCGA ACTGGATCTC AACAGCGGTA AGATCCTTGA

3251 GAGTTTTTCGC CCCGAAGAAC GTTTTCCAAT GATGAGCACT TTAAAGTTC
 3301 TGCTATGTGG CGCGGTATTA TCCCGTATTG ACGCCGGGCA AGAGCAACTC
 3351 GGTCGCCGCA TACACTATTC TCAGAATGAC TTGGTTGAGT ACTCACCAGT
 3401 CACAGAAAAG CATCTTACGG ATGGCATGAC AGTAAGAGAA TTATGCAGTG
 3451 CTGCCATAAC CATGAGTGAT AACACTGCGG CCAACTTACT TCTGACAACG
 3501 ATCGGAGGAC CGAAGGAGCT AACCGCTTTT TTGCACAACA TGGGGGATCA
 3551 TGTAACTCGC CTTGATCGTT GGAACCGGA GCTGAATGAA GCCATACCAA
 3601 ACGACGAGCG TGACACCACG ATGCCTGTAG CAATGGCAAC AACGTTGCGC
 3651 AACTATTAA CTGGCGAACT ACTTACTCTA GCTTCCCGGC AACAAATTAAT
 3701 AGACTGGATG GAGGCGGATA AAGTTGCAGG ACCACTTCTG CGCTCGGCCC
 3751 TTCCGGCTGG CTGGTTTATT GCTGATAAAT CTGGAGCCGG TGAGCGTGGG
 3801 TCTCGCGGTA TCATTGCAGC ACTGGGGCCA GATGGTAAGC CCTCCCGTAT
 3851 CGTAGTTATC TACACGACGG GGAGTCAGGC AACTATGGAT GAACGAAATA
 3901 GACAGATCGC TGAGATAGGT GCCTCACTGA TTAAGCATTG GTAAGTGTCA
 3951 GACCAAGTTT ACTCATATAT ACTTTAGATT GATTTAAAC TTCATTTTAA
 4001 ATTTAAAAGG ATCTAGGTGA AGATCCTTTT TGATAATCTC ATGACCAAAA
 4051 TCCCTTAACG TGAGTTTTTCG TTCCACTGAG CGTCAGACCC CGTAGAAAAG
 4101 ATCAAAGGAT CTTCTTGAGA TCCTTTTTTT CTGCGCGTAA TCTGCTGCTT
 4151 GCAAACAAAA AAACCACCGC TACCAGCGGT GGTTTGTTTG CCGGATCAAG
 4201 AGCTACCAAC TCTTTTCCG AAGGTAAGT GCTTCAGCAG AGCGCAGATA
 4251 CCAATACTG TCCTTCTAGT GTAGCCGTAG TTAGGCCACC ACTTCAAGAA

4301 CTCTGTAGCA CCGCCTACAT ACCTCGCTCT GCTAATCCTG TTACCAGTGG
 4351 CTGCTGCCAG TGGCGATAAG TCGTGTCTTA CCGGGTTGGA CTCAAGACGA
 4401 TAGTTACCGG ATAAGGCGCA GCGGTCGGGC TGAACGGGGG GTTCGTGCAC
 4451 ACAGCCCAGC TTGGAGCGAA CGACCTACAC CGAACTGAGA TACCTACAGC
 4501 GTGAGCTATG AGAAAGCGCC ACGCTTCCCG AAGGGAGAAA GGCGGACAGG
 4551 TATCCGGTAA GCGGCAGGGT CGGAACAGGA GAGCGCACGA GGGAGCTTCC
 4601 AGGGGGAAAC GCCTGGTATC TTTATAGTCC TGTCGGGTTT CGCCACCTCT
 4651 GACTTGAGCG TCGATTTTGG TGATGCTCGT CAGGGGGGCG GAGCCTATGG
 4701 AAAAACGCCA GCAACGCGGC CTTTTTACGG TTCCTGGCCT TTTGCTGGCC
 4751 TTTTGCTCAC ATGTTCTTTC CTGCGTTATC CCCTGATTCT GTGGATAACC
 4801 GTATTACCGC CTTTGAGTGA GCTGATACCG CTCGCCGAG CCGAACGACC
 4851 GAGCGCAGCG AGTCAGTGAG CGAGGAAGCG GAAGA